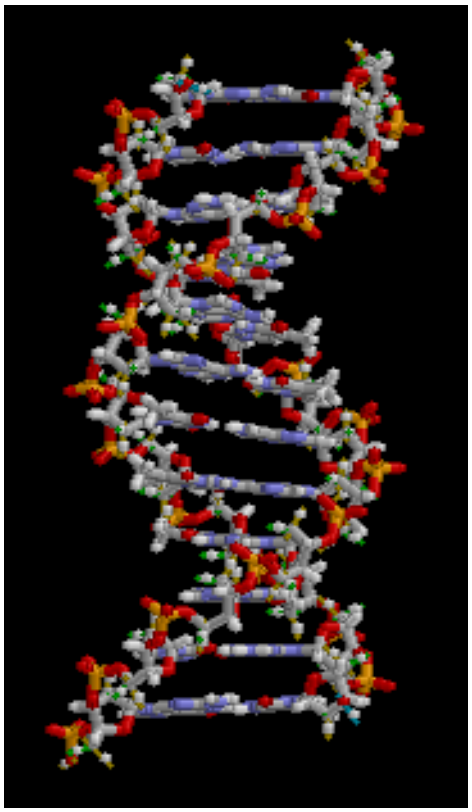
Interactomics

DNA,Proteomics,Metabolic Networks,Interactomics and Complex Systems Biology

Genomics

DNA

Deoxyribonucleic acid (**DNA**) is a nucleic acid that contains the genetic instructions used in the development and functioning of all known living organisms and some viruses. The main role of DNA molecules is the long-term storage of information. DNA is often compared to a set of blueprints or a recipe, or a code, since it contains the instructions needed to construct other components of cells, such as proteins and RNA molecules. The DNA segments that carry this genetic information are called genes, but other DNA sequences have structural purposes, or are involved in regulating the use of this genetic information.



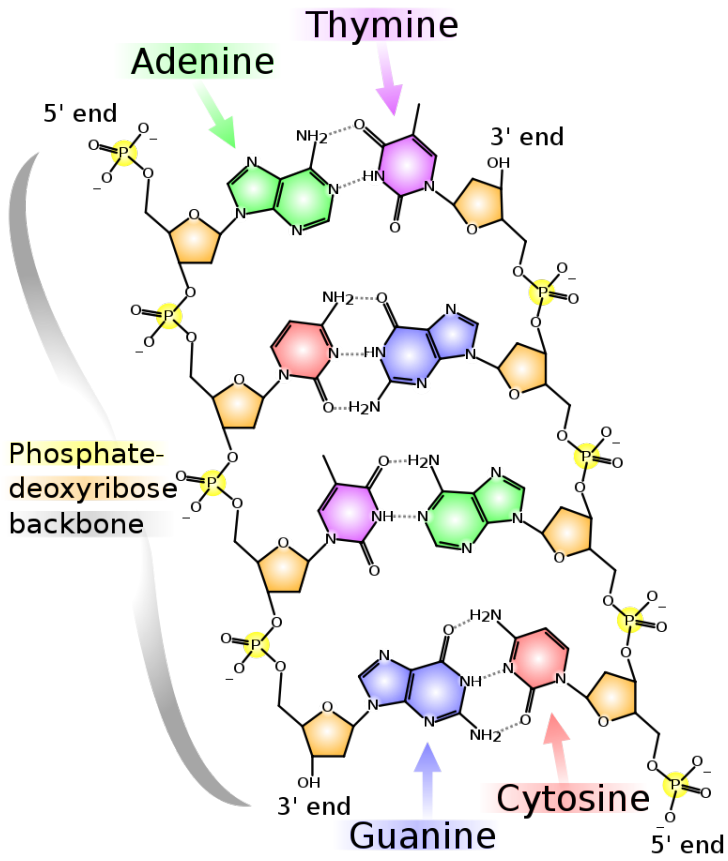
The structure of part of a DNA double helix

Chemically, DNA consists of two long polymers of simple units called nucleotides, with backbones made of sugars and phosphate groups joined by ester bonds. These two strands run in opposite directions to each other and are therefore anti-parallel. Attached to each sugar is one of four types of molecules called bases. It is the sequence of these four bases along the backbone that encodes information. This information is read using the genetic code, which specifies the sequence of the amino acids within proteins. The code is read by copying stretches of DNA into the related nucleic acid RNA, in a process called transcription.

Within cells, DNA is organized into structures called chromosomes. These chromosomes are duplicated before cells divide, in a process called DNA replication. Eukaryotic organisms (animals, plants, fungi, and protists) store most of their DNA inside the cell nucleus and some of their DNA in the mitochondria. Prokaryotes (bacteria and archaea) however, store their DNA in the cell's cytoplasm. Within the chromosomes, chromatin proteins such as histones compact and organize DNA. These compact structures guide the interactions between DNA and other proteins, helping control which parts of the DNA are transcribed.

Properties

DNA is a long polymer made from repeating units called nucleotides.[1] [2] [3] The DNA chain is 22 to 26 Ångströms wide (2.2 to 2.6 nanometres), and one nucleotide unit is 3.3 Å (0.33 nm) long.[4] Although each individual repeating unit is very small, DNA polymers can be very large molecules containing millions of nucleotides. For instance, the largest human chromosome, chromosome number 1, is approximately 220 million base pairs long.[5]



The chemical structure of DNA. Hydrogen bonds are shown as dotted lines.

In living organisms, DNA does not usually exist as a single molecule, but instead as a pair of molecules that are held tightly together.[6] [7] These two long strands entwine like vines, in the shape of a double helix. The nucleotide repeats contain both the segment of the backbone of the molecule, which holds the chain together, and a base, which interacts with the other DNA strand in the helix. In general, a base linked to a sugar is called a nucleoside and a base linked to a sugar and one or more phosphate groups is called a nucleotide. If multiple nucleotides are linked together, as in DNA, this polymer is called a polynucleotide.[8]

The backbone of the DNA strand is made from alternating phosphate and sugar residues.[9] The sugar in DNA is 2-deoxyribose, which is a pentose (five-carbon) sugar. The sugars are joined together by phosphate groups that form phosphodiester bonds between the third and fifth carbon atoms of adjacent sugar rings. These asymmetric bonds mean a strand of DNA has a direction. In a double helix the direction of the nucleotides in one strand is opposite to their direction in the other strand. This arrangement of DNA strands is called antiparallel. The asymmetric ends of DNA strands are referred to as the 5′ (*five prime*) and 3′ (*three prime*) ends, with the 5' end being that with a terminal phosphate group and the 3' end that with a terminal hydroxyl group. One of the major differences between DNA and RNA is the sugar, with 2-deoxyribose being replaced by the alternative pentose sugar ribose in RNA.[7]

The DNA double helix is stabilized by hydrogen bonds between the bases attached to the two strands. The four bases found in DNA are adenine (abbreviated A), cytosine (C), guanine (G) and thymine (T). These four bases are attached to the sugar/phosphate to form the complete nucleotide, as shown for adenosine monophosphate.

These bases are classified into two types; adenine and guanine are fused five- and six-membered heterocyclic compounds called purines, while cytosine and thymine are six-membered rings called pyrimidines.[7] A fifth pyrimidine base, called uracil (U), usually takes the place of thymine in RNA and differs from thymine by lacking a methyl group on its ring. Uracil is not usually found in DNA, occurring only as a breakdown product of cytosine.

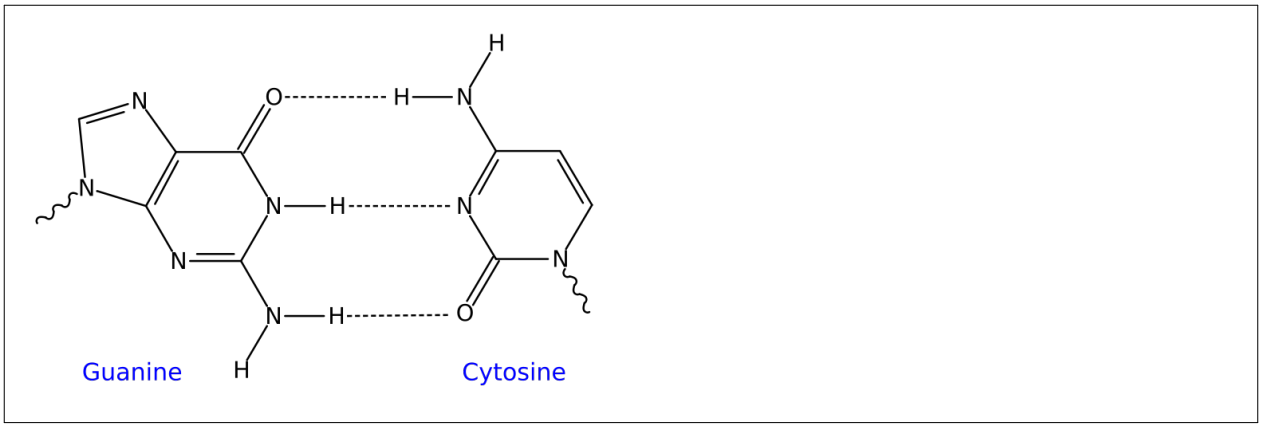
Grooves

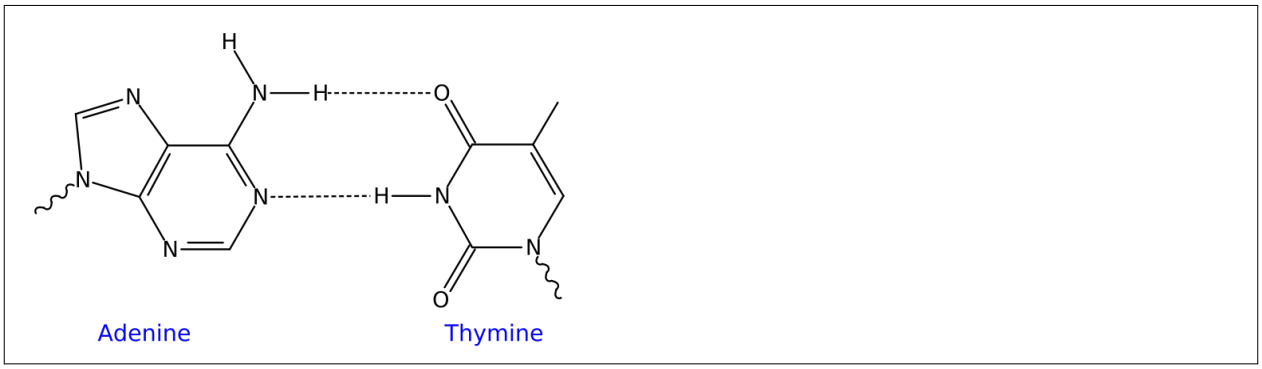
Twin helical strands form the DNA backbone. Another double helix may be found by tracing the spaces, or grooves, between the strands. These voids are adjacent to the base pairs and may provide a binding site. As the strands are not directly opposite each other, the grooves are unequally sized. One groove, the major groove, is 22 Å wide and the other, the minor groove, is 12 Å wide.[11] The narrowness of the minor groove means that the edges of the bases are more accessible in the major groove. As a result, proteins like transcription factors that can bind to specific sequences in double-stranded DNA usually make contacts to the sides of the bases exposed in the major groove.[12] This situation varies in unusual conformations of DNA within the cell *(see below)*, but the major and minor grooves are always named to reflect the differences in size that would be seen if the DNA is twisted back into the ordinary B form.

Structure of a section of DNA. The bases lie horizontally between the two spiraling strands.[10] Animated version at File:DNA orbit animated.gif - over 3 megabytes.

Base pairing

Each type of base on one strand forms a bond with just one type of base on the other strand. This is called complementary base pairing. Here, purines form hydrogen bonds to pyrimidines, with A bonding only to T, and C bonding only to G. This arrangement of two nucleotides binding together across the double helix is called a base pair. As hydrogen bonds are not covalent, they can be broken and rejoined relatively easily. The two strands of DNA in a double helix can therefore be pulled apart like a zipper, either by a mechanical force or high temperature.[13] As a result of this complementarity, all the information in the double-stranded sequence of a DNA helix is duplicated on each strand, which is vital in DNA replication. Indeed, this reversible and specific interaction between complementary base pairs is critical for all the functions of DNA in living organisms.[2]





Top, a **GC** base pair with three hydrogen bonds. Bottom, an **AT** base pair with two hydrogen bonds. Non-covalent hydrogen bonds between the pairs are shown as dashed lines.

The two types of base pairs form different numbers of hydrogen bonds, AT forming two hydrogen bonds, and GC forming three hydrogen bonds (see figures, left). DNA with high GC-content is more stable than DNA with low GC-content, but contrary to popular belief, this is not due to the extra hydrogen bond of a GC basepair but rather the contribution of stacking interactions (hydrogen bonding merely provides specificity of the pairing, not stability).[14] As a result, it is both the percentage of GC base pairs and the overall length of a DNA double helix that determine the strength of the association between the two strands of DNA. Long DNA helices with a high GC content have stronger-interacting strands, while short helices with high AT content have weaker-interacting strands.[15] In biology, parts of the DNA double helix that need to separate easily, such as the TATAAT Pribnow box in some promoters, tend to have a high AT content, making the strands easier to pull apart.[16] In the laboratory, the strength of this interaction can be measured by finding the temperature required to break the hydrogen bonds, their melting temperature (also called T_m value). When all the base pairs in a DNA double helix melt, the strands separate and exist in solution as two entirely independent molecules. These single-stranded DNA molecules have no single common shape, but some conformations are more stable than others.[17]

Sense and antisense

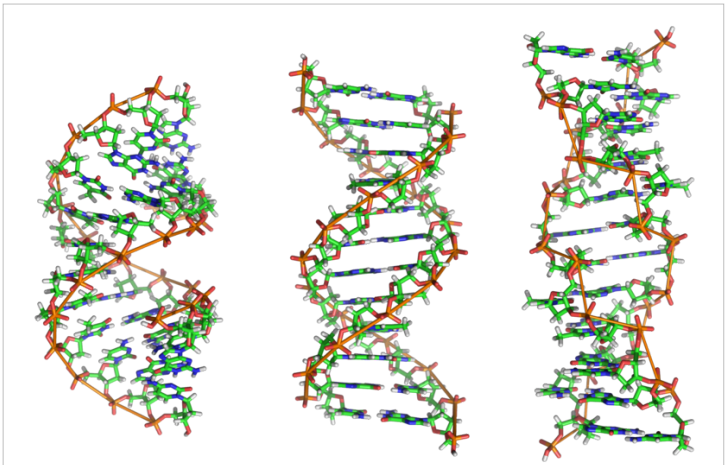
A DNA sequence is called "sense" if its sequence is the same as that of a messenger RNA copy that is translated into protein.[18] The sequence on the opposite strand is called the "antisense" sequence. Both sense and antisense sequences can exist on different parts of the same strand of DNA (i.e. both strands contain both sense and antisense sequences). In both prokaryotes and eukaryotes, antisense RNA sequences are produced, but the functions of these RNAs are not entirely clear.[19] One proposal is that antisense RNAs are involved in regulating gene expression through RNA-RNA base pairing.[20]

A few DNA sequences in prokaryotes and eukaryotes, and more in plasmids and viruses, blur the distinction between sense and antisense strands by having overlapping genes.[21] In these cases, some DNA sequences do double duty, encoding one protein when read along one strand, and a second protein when read in the opposite direction along the other strand. In bacteria, this overlap may be involved in the regulation of gene transcription,[22] while in viruses, overlapping genes increase the amount of information that can be encoded within the small viral genome.[23]

Supercoiling

DNA can be twisted like a rope in a process called DNA supercoiling. With DNA in its "relaxed" state, a strand usually circles the axis of the double helix once every 10.4 base pairs, but if the DNA is twisted the strands become more tightly or more loosely wound.[24] If the DNA is twisted in the direction of the helix, this is positive supercoiling, and the bases are held more tightly together. If they are twisted in the opposite direction, this is negative supercoiling, and the bases come apart more easily. In nature, most DNA has slight negative supercoiling that is introduced by enzymes called topoisomerases.[25] These enzymes are also needed to relieve the twisting stresses introduced into DNA strands during processes such as transcription and DNA replication.[26]

Alternate DNA structures



From left to right, the structures of A, B and Z DNA

DNA exists in many possible conformations that include A-DNA, B-DNA, and Z-DNA forms, although, only B-DNA and Z-DNA have been directly observed in functional organisms.[9] The conformation that DNA adopts depends on the hydration level, DNA sequence, the amount and direction of supercoiling, chemical modifications of the bases, the type and concentration of metal ions, as well as the presence of polyamines in solution.[27]

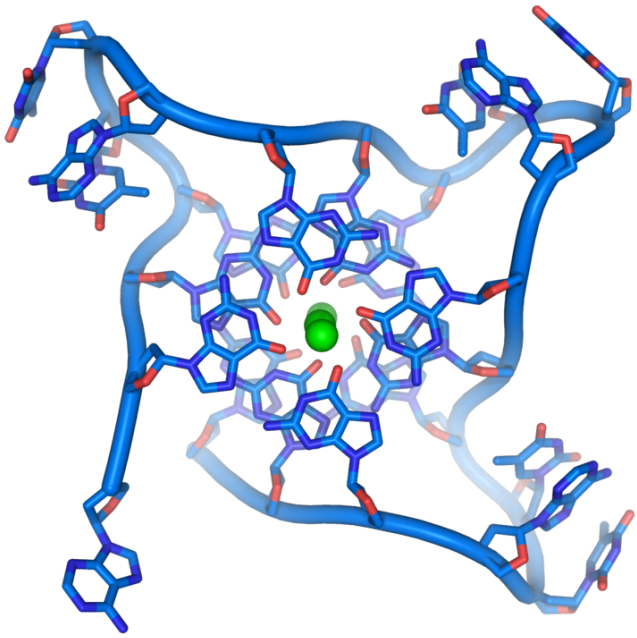
The first published reports of A-DNA X-ray diffraction patterns— and also B-DNA used analyses based on Patterson transforms that provided only a limited amount of structural information for oriented fibers of DNA.[28] [29] An alternate analysis was then proposed by Wilkins *et al.*, in 1953, for the *in vivo* B-DNA X-ray diffraction/scattering patterns of highly

hydrated DNA fibers in terms of squares of Bessel functions.[30] In the same journal, Watson and Crick presented their → molecular modeling analysis of the DNA X-ray diffraction patterns to suggest that the structure was a double-helix.[6]

Although the `B-DNA form' is most common under the conditions found in cells,[31] it is not a well-defined conformation but a family of related DNA conformations[32] that occur at the high hydration levels present in living cells. Their corresponding X-ray diffraction and scattering patterns are characteristic of molecular paracrystals with a significant degree of disorder.[33] [34]

Compared to B-DNA, the A-DNA form is a wider right-handed spiral, with a shallow, wide minor groove and a narrower, deeper major groove. The A form occurs under non-physiological conditions in partially dehydrated samples of DNA, while in the cell it may be produced in hybrid pairings of DNA and RNA strands, as well as in enzyme-DNA complexes.[35] [36] Segments of DNA where the bases have been chemically modified by methylation may undergo a larger change in conformation and adopt the Z form. Here, the strands turn about the helical axis in a left-handed spiral, the opposite of the more common B form.[37] These unusual structures can be recognized by specific Z-DNA binding proteins and may be involved in the regulation of transcription.[38]

Quadruplex structures



Structure of a DNA quadruplex formed by telomere repeats. The looped conformation of the DNA backbone is very different from the typical helical structure.[39]

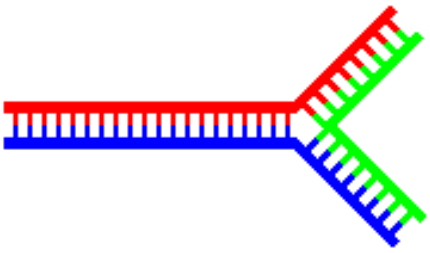
At the ends of the linear chromosomes are specialized regions of DNA called telomeres. The main function of these regions is to allow the cell to replicate chromosome ends using the enzyme telomerase, as the enzymes that normally replicate DNA cannot copy the extreme 3′ ends of chromosomes.[40] These specialized chromosome caps also help protect the DNA ends, and stop the DNA repair systems in the cell from treating them as damage to be corrected.[41] In human cells, telomeres are usually lengths of single-stranded DNA containing several thousand repeats of a simple TTAGGG sequence.[42]

These guanine-rich sequences may stabilize chromosome ends by forming structures of stacked sets of four-base units, rather than the usual base pairs found in other DNA molecules. Here, four guanine bases form a flat plate and these flat four-base units then stack on top of each other, to form a stable *G-quadruplex* structure.[43] These structures are stabilized by hydrogen bonding between the edges of the bases and chelation of a metal ion in the centre of each four-base unit.[44]

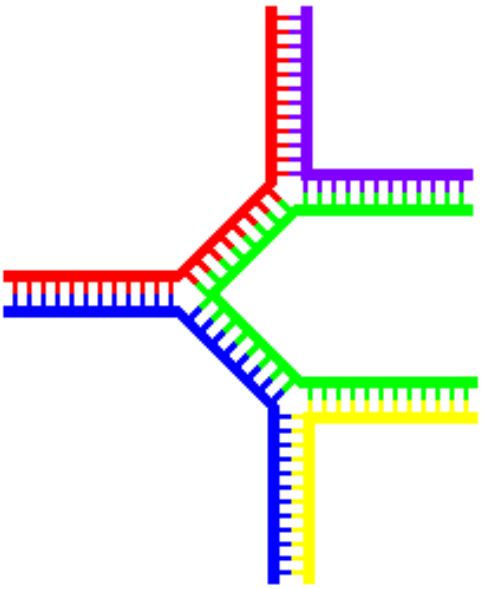
Other structures can also be formed, with the central set of four bases coming from either a single strand folded around the bases, or several different parallel strands, each contributing one base to the central structure.

In addition to these stacked structures, telomeres also form large loop structures called telomere loops, or T-loops. Here, the single-stranded DNA curls around in a long circle stabilized by telomere-binding proteins.[45] At the very end of the T-loop, the single-stranded telomere DNA is held onto a region of double-stranded DNA by the telomere strand disrupting the double-helical DNA and base pairing to one of the two strands. This triple-stranded structure is called a displacement loop or D-loop.[43]

Branched DNA



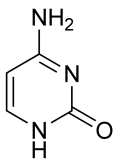
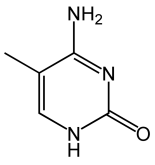
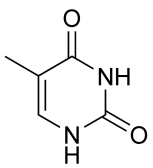
A DNA structure with a single branching point.



A DNA structure with multiple branches.

In DNA fraying occurs when non-complementary regions exist at the end of an otherwise complementary double-strand of DNA. However, branched DNA can occur if a third strand of DNA is introduced and contains adjoining regions able to hybridize with the frayed regions of the pre-existing double-strand. Although the simplest example of branched DNA involves only three strands of DNA, complexes involving additional strands and multiple branches are also possible.[46]

Chemical modifications

		
cytosine	5-methylcytosine	thymine

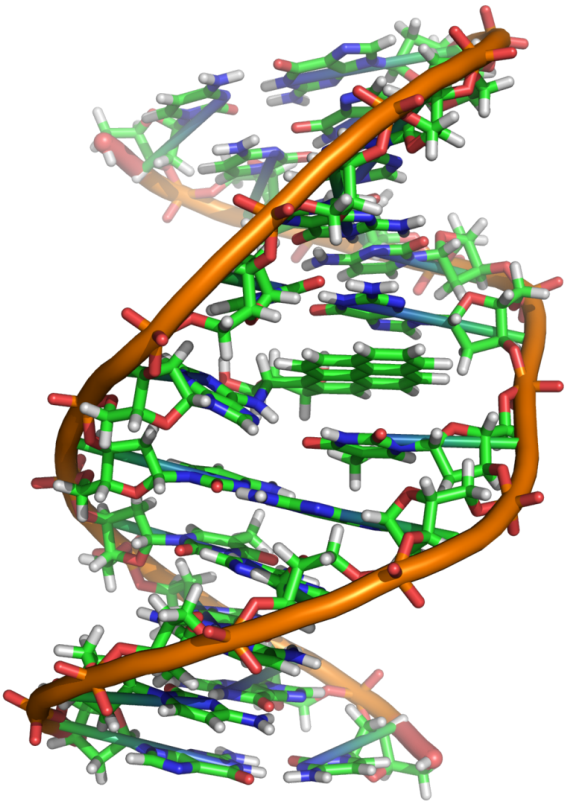
Structure of cytosine with and without the 5-methyl group. After deamination the 5-methylcytosine has the same structure as thymine

Base modifications

The expression of genes is influenced by how the DNA is packaged in chromosomes, in a structure called chromatin. Base modifications can be involved in packaging, with regions that have low or no gene expression usually containing high levels of methylation of cytosine bases. For example, cytosine methylation, produces 5-methylcytosine, which is important for X-chromosome inactivation.[47] The average level of methylation varies between organisms - the worm *Caenorhabditis elegans* lacks cytosine methylation, while vertebrates have higher levels, with up to 1% of their DNA containing 5-methylcytosine.[48] Despite the importance of 5-methylcytosine, it can deaminate to leave a thymine base, methylated cytosines are therefore particularly prone to mutations.[49] Other base modifications include adenine methylation in bacteria, the presence of 5-hydroxymethylcytosine in the brain,[50] and the glycosylation of uracil to produce the "J-base" in kinetoplastids.[51] [52]

Damage

DNA can be damaged by many different sorts of mutagens, which change the DNA sequence. Mutagens include oxidizing agents, alkylating agents and also high-energy electromagnetic radiation such as ultraviolet light and X-rays. The type of DNA damage produced depends on the type of mutagen. For example, UV light can damage DNA by producing thymine dimers, which are cross-links between pyrimidine bases.[54] On the other hand, oxidants such as free radicals or hydrogen peroxide produce multiple forms of damage, including base modifications, particularly of guanosine, and double-strand breaks.[55] A typical human cell contains about 150,000 bases that have suffered oxidative damage.[56] Of these oxidative lesions, the most dangerous are double-strand breaks, as these are difficult to repair and can produce point mutations, insertions and deletions from the DNA sequence, as well as chromosomal translocations.[57]



A covalent adduct between benzo[*a*]pyrene, the major mutagen in tobacco smoke, and DNA[53]

Many mutagens fit into the space between two adjacent base pairs, this is called *intercalating*. Most intercalators are aromatic and planar molecules, and include Ethidium bromide, daunomycin, and doxorubicin. In order for an intercalator to fit between base

pairs, the bases must separate, distorting the DNA strands by unwinding of the double helix. This inhibits both transcription and DNA replication, causing toxicity and mutations. As a result, DNA intercalators are often carcinogens, and Benzo[*a*]pyrene diol epoxide, acridines, aflatoxin and ethidium bromide are well-known examples.[58] [59] [60] Nevertheless, due to their ability to inhibit DNA transcription and replication, other similar toxins are also used in chemotherapy to inhibit rapidly growing cancer cells.[61]

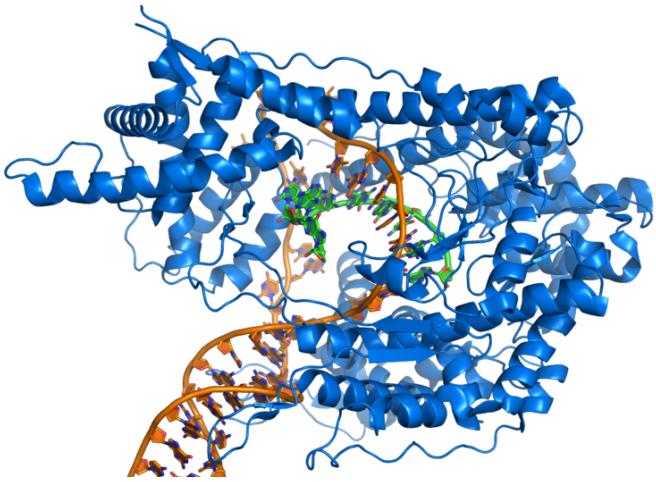
Biological functions

DNA usually occurs as linear chromosomes in eukaryotes, and circular chromosomes in prokaryotes. The set of chromosomes in a cell makes up its genome; the human genome has approximately 3 billion base pairs of DNA arranged into 46 chromosomes.[62] The information carried by DNA is held in the sequence of pieces of DNA called genes. Transmission of genetic information in genes is achieved via complementary base pairing. For example, in transcription, when a cell uses the information in a gene, the DNA sequence is copied into a complementary RNA sequence through the attraction between the DNA and the correct RNA nucleotides. Usually, this RNA copy is then used to make a matching protein sequence in a process called translation which depends on the same interaction between RNA nucleotides. Alternatively, a cell may simply copy its genetic information in a process called DNA replication. The details of these functions are covered in other articles; here we focus on the interactions between DNA and other molecules that mediate the function of the genome.

Genes and genomes

Genomic DNA is located in the cell nucleus of eukaryotes, as well as small amounts in mitochondria and chloroplasts. In prokaryotes, the DNA is held within an irregularly shaped body in the cytoplasm called the nucleoid.[63] The genetic information in a genome is held within genes, and the complete set of this information in an organism is called its genotype. A gene is a unit of heredity and is a region of DNA that influences a particular characteristic in an organism. Genes contain an open reading frame that can be transcribed, as well as regulatory sequences such as promoters and enhancers, which control the transcription of the open reading frame.

In many species, only a small fraction of the total sequence of the genome encodes protein. For example, only about 1.5% of the human genome consists of protein-coding exons, with over 50% of human DNA consisting of non-coding repetitive sequences.[64] The reasons for the presence of so much non-coding DNA in eukaryotic genomes and the extraordinary differences in genome size, or *C-value*, among species represent a long-standing puzzle known as the "C-value enigma."[65] However, DNA sequences that do not code protein may still encode functional non-coding RNA molecules, which are involved in the regulation of gene expression.[66]



T7 RNA polymerase (blue) producing a mRNA (green) from a DNA template (orange).[67]

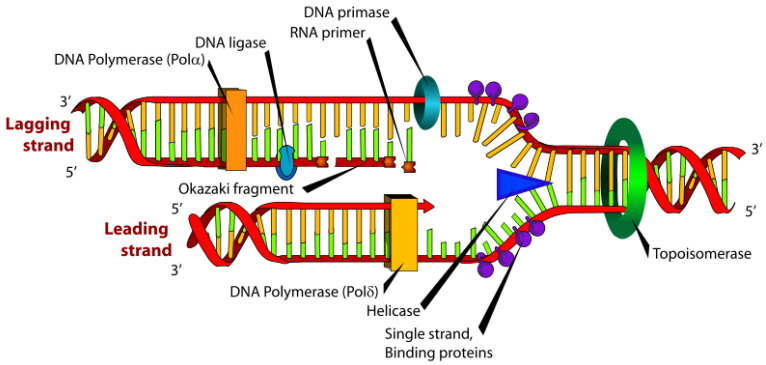
Some non-coding DNA sequences play structural roles in chromosomes. Telomeres and centromeres typically contain few genes, but are important for the function and stability of chromosomes.[41] [68] An abundant form of non-coding DNA in humans are pseudogenes, which are copies of genes that have been disabled by mutation.[69] These sequences are usually just molecular fossils, although they can occasionally serve as raw genetic material for the creation of new genes through the process of gene duplication and divergence.[70]

Transcription and translation

A gene is a sequence of DNA that contains genetic information and can influence the phenotype of an organism. Within a gene, the sequence of bases along a DNA strand defines a messenger RNA sequence, which then defines one or more protein sequences. The relationship between the nucleotide sequences of genes and the amino-acid sequences of proteins is determined by the rules of translation, known collectively as the genetic code. The genetic code consists of three-letter 'words' called *codons* formed from a sequence of three nucleotides (e.g. ACT, CAG, TTT).

In transcription, the codons of a gene are copied into messenger RNA by RNA polymerase. This RNA copy is then decoded by a ribosome that reads the RNA sequence by base-pairing the messenger RNA to transfer RNA, which carries amino acids. Since there are 4 bases in 3-letter combinations, there are 64 possible codons (4^3 combinations). These encode the twenty standard amino acids, giving most amino acids more than one possible codon. There are also three 'stop' or 'nonsense' codons signifying the end of the coding region; these are the TAA, TGA and TAG codons.

Replication



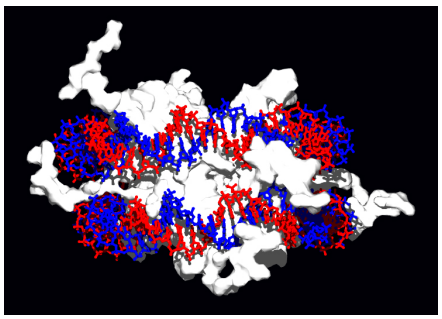
DNA replication. The double helix is unwound by a helicase and topoisomerase. Next, one DNA polymerase produces the leading strand copy. Another DNA polymerase binds to the lagging strand. This enzyme makes discontinuous segments (called Okazaki fragments) before DNA ligase joins them together.

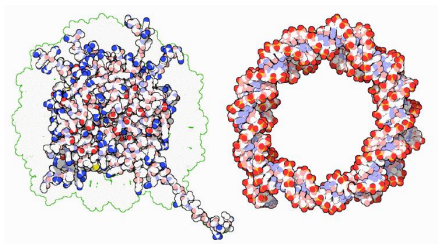
Cell division is essential for an organism to grow, but when a cell divides it must replicate the DNA in its genome so that the two daughter cells have the same genetic information as their parent. The double-stranded structure of DNA provides a simple mechanism for DNA replication. Here, the two strands are separated and then each strand's complementary DNA sequence is recreated by an enzyme called DNA polymerase. This enzyme makes the complementary strand by finding the correct base through complementary base pairing, and bonding it onto the original strand. As DNA polymerases can only extend a DNA strand in a 5′ to 3′ direction, different mechanisms are used to copy the antiparallel strands of the double helix.[71] In this way, the base on the old strand dictates which base appears on the new strand, and the cell ends up with a perfect copy of its DNA.

Interactions with proteins

All the functions of DNA depend on interactions with proteins. These protein interactions can be non-specific, or the protein can bind specifically to a single DNA sequence. Enzymes can also bind to DNA and of these, the polymerases that copy the DNA base sequence in transcription and DNA replication are particularly important.

DNA-binding proteins



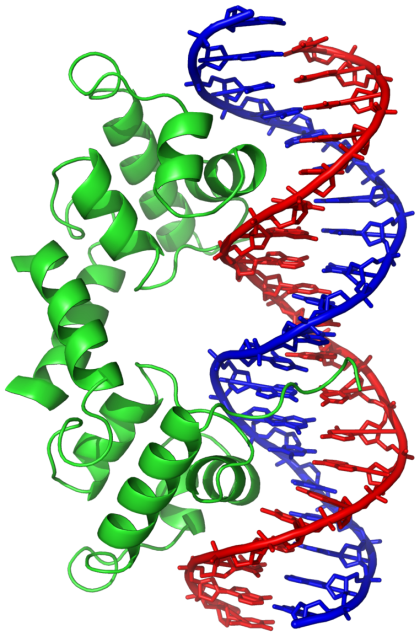


Interaction of DNA with histones (shown in white, top). These proteins' basic amino acids (below left, blue) bind to the acidic phosphate groups on DNA (below right, red).

Structural proteins that bind DNA are well-understood examples of non-specific DNA-protein interactions. Within chromosomes, DNA is held in complexes with structural proteins. These proteins organize the DNA into a compact structure called chromatin. In eukaryotes this structure involves DNA binding to a complex of small basic proteins called histones, while in prokaryotes multiple types of proteins are involved.[72] [73] The histones form a disk-shaped complex called a nucleosome, which contains two complete turns of double-stranded DNA wrapped around its surface. These non-specific interactions are formed through basic residues in the histones making ionic bonds to the acidic sugar-phosphate backbone of the DNA, and are therefore largely independent of the base sequence.[74] Chemical modifications of these basic amino acid residues include methylation, phosphorylation and acetylation.[75] These chemical changes alter the strength of the interaction between the DNA and the histones, making the DNA more or less accessible to transcription factors and changing the rate of transcription.[76] Other non-specific DNA-binding proteins in chromatin include the high-mobility group proteins, which bind to bent or distorted DNA.[77] These proteins are important in bending arrays of nucleosomes and arranging them into the larger structures that make up chromosomes.[78]

A distinct group of DNA-binding proteins are the DNA-binding proteins that specifically bind single-stranded DNA. In humans, replication protein A is the best-understood member of this family and is used in processes where the double helix is separated, including DNA replication, recombination and DNA repair.[79] These binding proteins seem to stabilize single-stranded DNA and protect it from forming stem-loops or being degraded by nucleases.

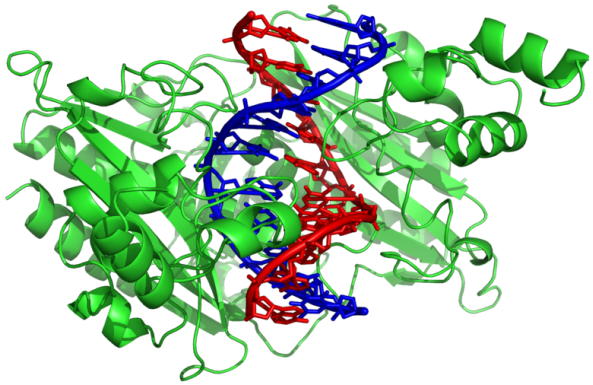
In contrast, other proteins have evolved to bind to particular DNA sequences. The most intensively studied of these are the various transcription factors, which are proteins that regulate transcription. Each transcription factor binds to one particular set of DNA sequences and activates or inhibits the transcription of genes that have these sequences close to their promoters. The transcription factors do this in two ways. Firstly, they can bind the RNA polymerase responsible for transcription, either directly or through other mediator proteins; this locates the polymerase at the promoter and allows it to begin transcription.[81] Alternatively, transcription factors can bind enzymes that modify the histones at the promoter; this will change the accessibility of the DNA template to the polymerase.[82]



The lambda repressor helix-turn-helix transcription factor bound to its DNA target[80]

As these DNA targets can occur throughout an organism's genome, changes in the activity of one type of transcription factor can affect thousands of genes.[83] Consequently, these proteins are often the targets of the signal transduction processes that control responses to environmental changes or cellular differentiation and development. The specificity of these transcription

factors' interactions with DNA come from the proteins making multiple contacts to the edges of the DNA bases, allowing them to "read" the DNA sequence. Most of these base-interactions are made in the major groove, where the bases are most accessible.[84]



The restriction enzyme EcoRV (green) in a complex with its substrate DNA[85]

DNA-modifying enzymes

Nucleases and ligases

Nucleases are enzymes that cut DNA strands by catalyzing the hydrolysis of the phosphodiester bonds. Nucleases that hydrolyse nucleotides from the ends of DNA strands are called exonucleases, while endonucleases cut within strands. The most frequently used nucleases in molecular biology are the restriction endonucleases, which cut DNA at specific sequences. For instance, the EcoRV enzyme shown to the left recognizes the 6-base sequence 5′-GAT|ATC-3′ and makes a cut at the vertical line. In nature, these enzymes protect bacteria against phage infection by digesting the phage DNA when it enters the bacterial cell, acting as part of the restriction modification system.[86] In technology, these sequence-specific nucleases are used in molecular cloning and DNA fingerprinting.

Enzymes called DNA ligases can rejoin cut or broken DNA strands.[87] Ligases are particularly important in lagging strand DNA replication, as they join together the short segments of DNA produced at the replication fork into a complete copy of the DNA template. They are also used in DNA repair and genetic recombination.[87]

Topoisomerases and helicases

Topoisomerases are enzymes with both nuclease and ligase activity. These proteins change the amount of supercoiling in DNA. Some of these enzyme work by cutting the DNA helix and allowing one section to rotate, thereby reducing its level of supercoiling; the enzyme then seals the DNA break.[25] Other types of these enzymes are capable of cutting one DNA helix and then passing a second strand of DNA through this break, before rejoining the helix.[88] Topoisomerases are required for many processes involving DNA, such as DNA replication and transcription.[26]

Helicases are proteins that are a type of molecular motor. They use the chemical energy in nucleoside triphosphates, predominantly ATP, to break hydrogen bonds between bases and unwind the DNA double helix into single strands.[89] These enzymes are essential for most processes where enzymes need to access the DNA bases.

Polymerases

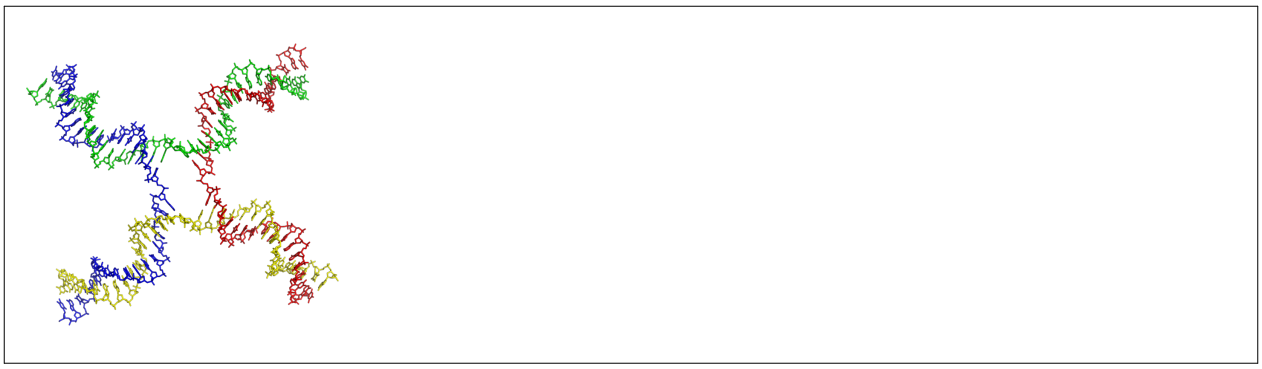
Polymerases are enzymes that synthesize polynucleotide chains from nucleoside triphosphates. The sequence of their products are copies of existing polynucleotide chains - which are called *templates*. These enzymes function by adding nucleotides onto the 3′ hydroxyl group of the previous nucleotide in a DNA strand. Consequently, all polymerases work in a 5′ to 3′ direction.[90] In the active site of these enzymes, the incoming nucleoside triphosphate base-pairs to the template: this allows polymerases to accurately synthesize the complementary strand of their template. Polymerases are classified according to the type of template that they use.

In DNA replication, a DNA-dependent DNA polymerase makes a copy of a DNA sequence. Accuracy is vital in this process, so many of these polymerases have a proofreading activity. Here, the polymerase recognizes the occasional mistakes in the synthesis reaction by the lack of base pairing between the mismatched nucleotides. If a mismatch is detected, a 3′ to 5′ exonuclease activity is activated and the incorrect base removed.[91] In most organisms DNA polymerases function in a large complex called the replisome that contains multiple accessory subunits, such as the DNA clamp or helicases.[92]

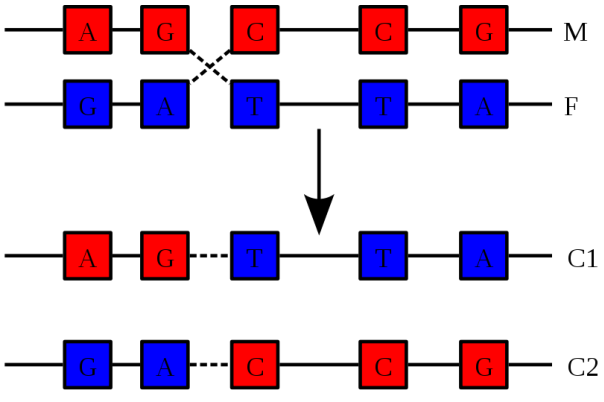
RNA-dependent DNA polymerases are a specialized class of polymerases that copy the sequence of an RNA strand into DNA. They include reverse transcriptase, which is a viral enzyme involved in the infection of cells by retroviruses, and telomerase, which is required for the replication of telomeres.[40] [93] Telomerase is an unusual polymerase because it contains its own RNA template as part of its structure.[41]

Transcription is carried out by a DNA-dependent RNA polymerase that copies the sequence of a DNA strand into RNA. To begin transcribing a gene, the RNA polymerase binds to a sequence of DNA called a promoter and separates the DNA strands. It then copies the gene sequence into a messenger RNA transcript until it reaches a region of DNA called the terminator, where it halts and detaches from the DNA. As with human DNA-dependent DNA polymerases, RNA polymerase II, the enzyme that transcribes most of the genes in the human genome, operates as part of a large protein complex with multiple regulatory and accessory subunits.[94]

Genetic recombination



Structure of the Holliday junction intermediate in genetic recombination. The four separate DNA strands are coloured red, blue, green and yellow.[95]



Recombination involves the breakage and rejoining of two chromosomes (M and F) to produce two re-arranged chromosomes (C1 and C2).

A DNA helix usually does not interact with other segments of DNA, and in human cells the different chromosomes even occupy separate areas in the nucleus called "chromosome territories".[96] This physical separation of different chromosomes is important for the ability of DNA to function as a stable repository for information, as one of the few times chromosomes interact is during chromosomal crossover when they recombine. Chromosomal crossover is when two DNA helices break, swap a section and then rejoin.

Recombination allows chromosomes to exchange genetic information and produces new combinations of genes, which increases the efficiency of natural selection and can be important in the rapid evolution of new proteins.[97] Genetic recombination can also be involved in DNA repair, particularly in the cell's response to double-strand breaks.[98]

The most common form of chromosomal crossover is homologous recombination, where the two chromosomes involved share very similar sequences. Non-homologous recombination can be damaging to cells, as it can produce chromosomal translocations and genetic abnormalities. The recombination reaction is catalyzed by enzymes known as *recombinases*, such as RAD51.[99] The first step in recombination is a double-stranded break either caused by an endonuclease or damage to the DNA.[100] A series of steps catalyzed in part by the recombinase then leads to joining of the two helices by at least one Holliday junction, in which a segment of a single strand in each helix is annealed to the complementary strand in the other helix. The Holliday junction is a tetrahedral junction structure that can be moved along the pair of chromosomes, swapping one strand for another. The recombination reaction is then halted by cleavage of the junction and re-ligation of the released DNA.[101]

Evolution

DNA contains the genetic information that allows all modern living things to function, grow and reproduce. However, it is unclear how long in the 4-billion-year history of life DNA has performed this function, as it has been proposed that the earliest forms of life may have used RNA as their genetic material.[90] [102] RNA may have acted as the central part of early cell metabolism as it can both transmit genetic information and carry out catalysis as part of ribozymes.[103] This ancient RNA world where nucleic acid would have been used for both catalysis and genetics may have influenced the evolution of the current genetic code based on four nucleotide bases. This would occur since the number of unique bases in such an organism is a trade-off between a small number of bases increasing replication accuracy and a large number of bases increasing the catalytic efficiency of ribozymes.[104]

Unfortunately, there is no direct evidence of ancient genetic systems, as recovery of DNA from most fossils is impossible. This is because DNA will survive in the environment for less than one million years and slowly degrades into short fragments in solution.[105] Claims for older DNA have been made, most notably a report of the isolation of a viable bacterium from a salt crystal 250-million years old,[106] but these claims are controversial.[107] [108]

Uses in technology

Genetic engineering

Methods have been developed to purify DNA from organisms, such as phenol-chloroform extraction and manipulate it in the laboratory, such as restriction digests and the polymerase chain reaction. Modern biology and biochemistry make intensive use of these techniques in recombinant DNA technology. Recombinant DNA is a man-made DNA sequence that has been assembled from other DNA sequences. They can be transformed into organisms in the form of plasmids or in the appropriate format, by using a viral vector.[109] The genetically modified organisms produced can be used to produce products such as recombinant proteins, used in medical research,[110] or be grown in agriculture.[111] [112]

Forensics

Forensic scientists can use DNA in blood, semen, skin, saliva or hair found at a crime scene to identify a matching DNA of an individual, such as a perpetrator. This process is called genetic fingerprinting, or more accurately, DNA profiling. In DNA profiling, the lengths of variable sections of repetitive DNA, such as short tandem repeats and minisatellites, are compared between people. This method is usually an extremely reliable technique for identifying a matching DNA.[113] However, identification can be complicated if the scene is contaminated with DNA from several people.[114] DNA profiling was developed in 1984 by British geneticist Sir Alec Jeffreys,[115] and first used in forensic science to convict Colin Pitchfork in the 1988 Enderby murders case.[116]

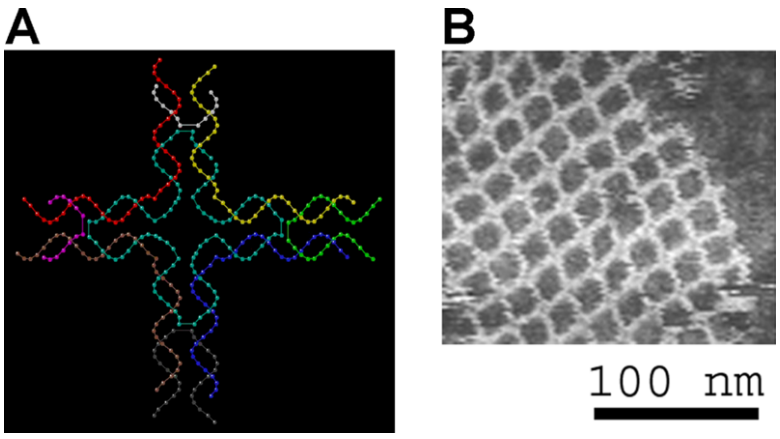
People convicted of certain types of crimes may be required to provide a sample of DNA for a database. This has helped investigators solve old cases where only a DNA sample was obtained from the scene. DNA profiling can also be used to identify victims of mass casualty incidents.[117] On the other hand, many convicted people have been released from prison on the basis of DNA techniques, which were not available when a crime had originally been committed.

Bioinformatics

→ Bioinformatics involves the manipulation, searching, and data mining of DNA sequence data. The development of techniques to store and search DNA sequences have led to widely applied advances in computer science, especially string searching algorithms, machine learning and database theory.[118] String searching or matching algorithms, which find an occurrence of a sequence of letters inside a larger sequence of letters, were developed to search for specific sequences of nucleotides.[119] In other applications such as text editors, even simple algorithms for this problem usually suffice, but DNA sequences cause these algorithms to exhibit near-worst-case behaviour due to their small number of distinct characters. The related problem of sequence alignment aims to identify homologous sequences and locate the specific mutations that make them distinct. These techniques, especially multiple sequence alignment, are used in studying phylogenetic relationships and protein function.[120] Data sets representing entire genomes' worth of DNA sequences, such as those produced by the Human Genome Project, are difficult to use without annotations, which label the locations of genes and regulatory elements on each chromosome. Regions of DNA sequence that have the characteristic patterns associated with protein- or RNA-coding genes can be identified by gene finding algorithms, which allow researchers to predict the presence of particular gene products in an organism even before they have been isolated experimentally.[121]

DNA nanotechnology

DNA nanotechnology uses the unique molecular recognition properties of DNA and other nucleic acids to create self-assembling branched DNA complexes with useful properties.[123] DNA is thus used as a structural material rather than as a carrier of biological information. This has led to the creation of two-dimensional periodic lattices (both tile-based as well as using the "DNA origami" method) as well as three-dimensional structures in the shapes of polyhedra.[124] Nanomechanical devices and algorithmic self-assembly have also been demonstrated,[125] and these DNA structures have been used to template the arrangement of other molecules such as gold nanoparticles and streptavidin proteins.[126]



The DNA structure at left (schematic shown) will self-assemble into the structure visualized by atomic force microscopy at right. DNA nanotechnology is the field which seeks to design nanoscale structures using the molecular recognition properties of DNA molecules. Image from Strong, 2004.[122]

History and anthropology

Because DNA collects mutations over time, which are then inherited, it contains historical information and by comparing DNA sequences, geneticists can infer the evolutionary history of organisms, their phylogeny.[127] This field of phylogenetics is a powerful tool in

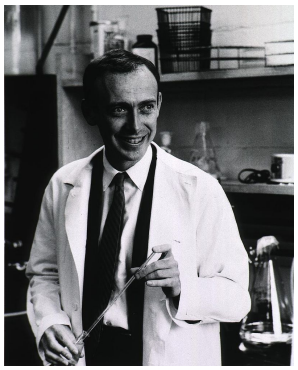
evolutionary biology. If DNA sequences within a species are compared, population geneticists can learn the history of particular populations. This can be used in studies ranging from ecological genetics to anthropology; for example, DNA evidence is being used to try to identify the Ten Lost Tribes of Israel.[128] [129]

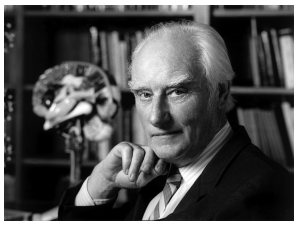
DNA has also been used to look at modern family relationships, such as establishing family relationships between the descendants of Sally Hemings and Thomas Jefferson. This usage is closely related to the use of DNA in criminal investigations detailed above. Indeed, some criminal investigations have been solved when DNA from crime scenes has matched relatives of the guilty individual.[130]

History of DNA research

DNA was first isolated by the Swiss physician Friedrich Miescher who, in 1869, discovered a microscopic substance in the pus of discarded surgical bandages. As it resided in the nuclei of cells, he called it "nuclein".[131] In 1919, Phoebus Levene identified the base, sugar and phosphate nucleotide unit.[132] Levene suggested that DNA consisted of a string of nucleotide units linked together through the phosphate groups. However, Levene thought the chain was short and the bases repeated in a fixed order. In 1937 William Astbury produced the first X-ray diffraction patterns that showed that DNA had a regular structure.[133]

In 1928, Frederick Griffith discovered that traits of the "smooth" form of the *Pneumococcus* could be transferred to the "rough" form of the same bacteria by mixing killed "smooth" bacteria with the live "rough" form.[134] This system provided the first clear suggestion that DNA carried genetic information—the Avery-MacLeod-McCarty experiment—when Oswald Avery, along with coworkers Colin MacLeod and Maclyn McCarty, identified DNA as the transforming principle in 1943.[135] DNA's role in heredity was confirmed in 1952, when Alfred Hershey and Martha Chase in the Hershey-Chase experiment showed that DNA is the genetic material of the T2 phage.[136]


James D. Watson

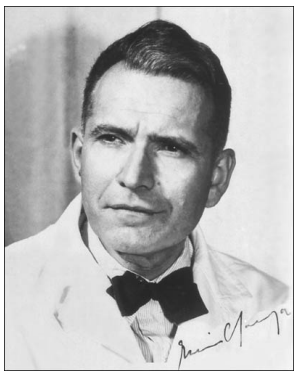

Francis Crick


Francis Crick

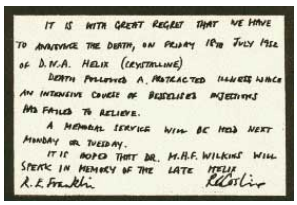

Rosalind Franklin

Raymond Gosling

Maurice F.Wilkins



Erwin Chargaff



DNA Helix controversy

In 1953 James D. Watson and Francis Crick suggested what is now accepted as the first correct double-helix model of DNA structure in the journal *Nature*.[6] Their double-helix, molecular model of DNA was then based on a single X-ray diffraction image (labeled as "Photo 51")[137] taken by Rosalind Franklin and Raymond Gosling in May 1952, as well as the information that the DNA bases were paired—also obtained through private communications from Erwin Chargaff in the previous years. Chargaff's rules played a very important role in establishing double-helix configurations for B-DNA as well as A-DNA.

Experimental evidence supporting the Watson and Crick model were published in a series of five articles in the same issue of *Nature*.[138] Of these, Franklin and Gosling's paper was the first publication of their own X-ray diffraction data and original analysis method that partially supported the Watson and Crick model[29] [139] ; this issue also contained an article on DNA structure by Maurice Wilkins and two of his colleagues, whose analysis and *in vivo* B-DNA X-ray patterns also supported the presence *in vivo* of the double-helical DNA configurations as proposed by Crick and Watson for their double-helix molecular model of DNA in the previous two pages of *Nature*.[30] In 1962, after Franklin's death, Watson, Crick, and Wilkins jointly received the Nobel Prize in Physiology or Medicine.[140] Unfortunately, Nobel rules of the time allowed only living recipients, but a vigorous debate continues on who should receive credit for the discovery.[141]

In an influential presentation in 1957, Crick laid out the "Central Dogma" of molecular biology, which foretold the relationship between DNA, RNA, and proteins, and articulated the "adaptor hypothesis".[142] Final confirmation of the replication mechanism that was implied by the double-helical structure followed in 1958 through the Meselson-Stahl experiment.[143] Further work by Crick and coworkers showed that the genetic code was based on non-overlapping triplets of bases, called codons, allowing Har Gobind Khorana, Robert W. Holley and Marshall Warren Nirenberg to decipher the genetic code.[144] These findings represent the birth of molecular biology.

See also

- Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid
- → Molecular models of DNA
- DNA microarray
- DNA sequencing
- Paracrystal model and theory
- X-ray scattering
- Crystallography
- X-ray crystallography
- Genetic disorder
- Junk DNA
- Nucleic acid analogues
- Nucleic acid methods
- Nucleic acid modeling
- Nucleic Acid Notations
- Phosphoramidite
- Plasmid
- Polymerase chain reaction
- *Proteopedia DNA* [145]
- Southern blot
- Triple-stranded DNA

Notes

[1] Saenger, Wolfram (1984). *Principles of Nucleic Acid Structure*. New York: Springer-Verlag. ISBN 0387907629.

[2] Alberts, Bruce; Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walters (2002). *Molecular Biology of the Cell; Fourth Edition (http://www.ncbi.nlm.nih.gov/books/bv.fcgi?call=bv.View..ShowTOC&rid=mboc4.TOC&depth=2)*. New York and London: Garland Science. ISBN 0-8153-3218-1. OCLC 145080076 48122761 57023651 69932405 (http://worldcat.org/oclc/145080076+48122761+57023651+69932405). .

[3] Butler, John M. (2001). *Forensic DNA Typing*. Elsevier. ISBN 978-0-12-147951-0. OCLC 223032110 45406517 (http://worldcat.org/oclc/223032110+45406517). pp. 14–15.

[4] Mandelkern M, Elias J, Eden D, Crothers D (1981). "The dimensions of DNA in solution". *J Mol Biol* **152** (1): 153–61. doi: 10.1016/0022-2836(81)90099-1 (http://dx.doi.org/10.1016/0022-2836(81)90099-1). PMID 7338906.

[5] Gregory S, *et al.* (2006). "The DNA sequence and biological annotation of human chromosome 1". *Nature* **441** (7091): 315–21. doi: 10.1038/nature04727 (http://dx.doi.org/10.1038/nature04727). PMID 16710414.

[6] Watson J.D. and Crick F.H.C. (1953). " A Structure for Deoxyribose Nucleic Acid (http://www.nature.com/nature/dna50/watsoncrick.pdf)" (PDF). *Nature* **171**: 737–738. doi: 10.1038/171737a0 (http://dx.doi.org/10.1038/171737a0). PMID 13054692. . Retrieved on 4 May 2009.

[7] Berg J., Tymoczko J. and Stryer L. (2002) *Biochemistry.* W. H. Freeman and Company ISBN 0-7167-4955-6

[8] Abbreviations and Symbols for Nucleic Acids, Polynucleotides and their Constituents (http://www.chem.qmul.ac.uk/iupac/misc/naabb.html) IUPAC-IUB Commission on Biochemical Nomenclature (CBN), Accessed 03 January 2006

[9] Ghosh A, Bansal M (2003). "A glossary of DNA structures from A to Z". *Acta Crystallogr D Biol Crystallogr* **59** (Pt 4): 620–6. doi: 10.1107/S0907444903003251 (http://dx.doi.org/10.1107/S0907444903003251). PMID 12657780.

[10] Created from PDB 1D65 (http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1D65)

[11] Wing R, Drew H, Takano T, Broka C, Tanaka S, Itakura K, Dickerson R (1980). "Crystal structure analysis of a complete turn of B-DNA". *Nature* **287** (5784): 755–8. doi: 10.1038/287755a0 (http://dx.doi.org/10.1038/287755a0). PMID 7432492.

[12] Pabo C, Sauer R (1984). "Protein-DNA recognition". *Annu Rev Biochem* **53**: 293–321. doi: 10.1146/annurev.bi.53.070184.001453 (http://dx.doi.org/10.1146/annurev.bi.53.070184.001453). PMID

6236744.

[13] Clausen-Schaumann H, Rief M, Tolksdorf C, Gaub H (2000). " Mechanical stability of single DNA molecules (http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez&artid=1300792)". *Biophys J* **78** (4): 1997–2007. doi: 10.1016/S0006-3495(00)76747-6 (http://dx.doi.org/10.1016/S0006-3495(00)76747-6). PMID 10733978.

[14] Yakovchuk P, Protozanova E, Frank-Kamenetskii MD (2006). " Base-stacking and base-pairing contributions into thermal stability of the DNA double helix (http://nar.oxfordjournals.org/cgi/pmidlookup?view=long&pmid=16449200)". *Nucleic Acids Res.* **34** (2): 564–74. doi: 10.1093/nar/gkj454 (http://dx.doi.org/10.1093/nar/gkj454). PMID 16449200. PMC: 1360284 (http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez&artid=1360284). .

[15] Chalikian T, Völker J, Plum G, Breslauer K (1999). " A more unified picture for the thermodynamics of nucleic acid duplex melting: a characterization by calorimetric and volumetric techniques (http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez&artid=22151)". *Proc Natl Acad Sci USA* **96** (14): 7853–8. doi: 10.1073/pnas.96.14.7853 (http://dx.doi.org/10.1073/pnas.96.14.7853). PMID 10393911.

[16] deHaseth P, Helmann J (1995). "Open complex formation by Escherichia coli RNA polymerase: the mechanism of polymerase-induced strand separation of double helical DNA". *Mol Microbiol* **16** (5): 817–24. doi: 10.1111/j.1365-2958.1995.tb02309.x (http://dx.doi.org/10.1111/j.1365-2958.1995.tb02309.x). PMID 7476180.

[17] Isaksson J, Acharya S, Barman J, Cheruku P, Chattopadhyaya J (2004). "Single-stranded adenine-rich DNA and RNA retain structural characteristics of their respective double-stranded conformations and show directional differences in stacking pattern". *Biochemistry* **43** (51): 15996–6010. doi: 10.1021/bi048221v (http://dx.doi.org/10.1021/bi048221v). PMID 15609994.

[18] Designation of the two strands of DNA (http://www.chem.qmul.ac.uk/iubmb/newsletter/misc/DNA.html) JCBN/NC-IUB Newsletter 1989, Accessed 07 May 2008

[19] Hüttenhofer A, Schattner P, Polacek N (2005). "Non-coding RNAs: hope or hype?". *Trends Genet* **21** (5): 289–97. doi: 10.1016/j.tig.2005.03.007 (http://dx.doi.org/10.1016/j.tig.2005.03.007). PMID 15851066.

[20] Munroe S (2004). "Diversity of antisense regulation in eukaryotes: multiple mechanisms, emerging patterns". *J Cell Biochem* **93** (4): 664–71. doi: 10.1002/jcb.20252 (http://dx.doi.org/10.1002/jcb.20252). PMID 15389973.

[21] Makalowska I, Lin C, Makalowski W (2005). "Overlapping genes in vertebrate genomes". *Comput Biol Chem* **29** (1): 1–12. doi: 10.1016/j.compbiolchem.2004.12.006 (http://dx.doi.org/10.1016/j.compbiolchem.2004.12.006). PMID 15680581.

[22] Johnson Z, Chisholm S (2004). "Properties of overlapping genes are conserved across microbial genomes". *Genome Res* **14** (11): 2268–72. doi: 10.1101/gr.2433104 (http://dx.doi.org/10.1101/gr.2433104). PMID 15520290.

[23] Lamb R, Horvath C (1991). "Diversity of coding strategies in influenza viruses". *Trends Genet* **7** (8): 261–6. PMID 1771674.

[24] Benham C, Mielke S (2005). "DNA mechanics". *Annu Rev Biomed Eng* **7**: 21–53. doi: 10.1146/annurev.bioeng.6.062403.132016 (http://dx.doi.org/10.1146/annurev.bioeng.6.062403.132016). PMID 16004565.

[25] Champoux J (2001). "DNA topoisomerases: structure, function, and mechanism". *Annu Rev Biochem* **70**: 369–413. doi: 10.1146/annurev.biochem.70.1.369 (http://dx.doi.org/10.1146/annurev.biochem.70.1.369). PMID 11395412.

[26] Wang J (2002). "Cellular roles of DNA topoisomerases: a molecular perspective". *Nat Rev Mol Cell Biol* **3** (6): 430–40. doi: 10.1038/nrm831 (http://dx.doi.org/10.1038/nrm831). PMID 12042765.

[27] Basu H, Feuerstein B, Zarling D, Shafer R, Marton L (1988). "Recognition of Z-RNA and Z-DNA determinants by polyamines in solution: experimental and theoretical studies". *J Biomol Struct Dyn* **6** (2): 299–309. PMID 2482766.

[28] Franklin RE, Gosling RG (6 March 1953). " The Structure of Sodium Thymonucleate Fibres I. The Influence of Water Content (http://hekto.med.unc.edu:8080/CARTER/carter_WWW/Bioch_134/PDF_files/Franklin_Gossling.pdf)". *Acta Cryst* **6** (8-9): 673–7. doi: 10.1107/S0365110X53001939 (http://dx.doi.org/10.1107/S0365110X53001939). .

Franklin RE, Gosling RG (September 1953). "The structure of sodium thymonucleate fibres. II. The cylindrically symmetrical Patterson function". *Acta Cryst* **6** (8-9): 678–85. doi: 10.1107/S0365110X53001940 (http://dx.doi.org/10.1107/S0365110X53001940).

[29] Franklin, Rosalind and Gosling, Raymond (1953). " Molecular Configuration in Sodium Thymonucleate. Franklin R. and Gosling R.G (http://www.nature.com/nature/dna50/franklingosling.pdf)" (PDF). *Nature* **171**: 740–1. doi: 10.1038/171740a0 (http://dx.doi.org/10.1038/171740a0). PMID 13054694. .

[30] Wilkins M.H.F., A.R. Stokes A.R. & Wilson, H.R. (1953). " Molecular Structure of Deoxypentose Nucleic Acids (http://www.nature.com/nature/dna50/wilkins.pdf)" (PDF). *Nature* **171**: 738–740. doi: 10.1038/171738a0 (http://dx.doi.org/10.1038/171738a0). PMID 13054693. .

[31] Leslie AG, Arnott S, Chandrasekaran R, Ratliff RL (1980). "Polymorphism of DNA double helices". *J. Mol. Biol.* **143** (1): 49–72. doi: 10.1016/0022-2836(80)90124-2 (http://dx.doi.org/10.1016/0022-2836(80)90124-2). PMID 7441761.

[32] Baianu, I.C. (1980). "Structural Order and Partial Disorder in Biological systems". *Bull. Math. Biol.* **42** (4): 137–141. http://cogprints.org/3822/

[33] Hosemann R., Bagchi R.N., *Direct analysis of diffraction by matter*, North-Holland Publs., Amsterdam – New York, 1962.

[34] Baianu, I.C. (1978). "X-ray scattering by partially disordered membrane systems.". *Acta Cryst.,* **A34** (5): 751–753. doi: 10.1107/S0567739478001540 (http://dx.doi.org/10.1107/S0567739478001540).

[35] Wahl M, Sundaralingam M (1997). "Crystal structures of A-DNA duplexes". *Biopolymers* **44** (1): 45–63. doi:10.1002/(SICI)1097-0282(1997)44:1 (inactive 2009-03-14) . PMID 9097733.

[36] Lu XJ, Shakked Z, Olson WK (2000). "A-form conformational motifs in ligand-bound DNA structures". *J. Mol. Biol.* **300** (4): 819–40. doi: 10.1006/jmbi.2000.3690 (http://dx.doi.org/10.1006/jmbi.2000.3690). PMID 10891271.

[37] Rothenburg S, Koch-Nolte F, Haag F (2001). "DNA methylation and Z-DNA formation as mediators of quantitative differences in the expression of alleles". *Immunol Rev* **184**: 286–98. doi: 10.1034/j.1600-065x.2001.1840125.x (http://dx.doi.org/10.1034/j.1600-065x.2001.1840125.x). PMID 12086319.

[38] Oh D, Kim Y, Rich A (2002). " Z-DNA-binding proteins can act as potent effectors of gene expression in vivo (http://www.pnas.org/cgi/pmidlookup?view=long&pmid=12486233)". *Proc. Natl. Acad. Sci. U.S.A.* **99** (26): 16666–71. doi: 10.1073/pnas.262672699 (http://dx.doi.org/10.1073/pnas.262672699). PMID 12486233. PMC: 139201 (http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez&artid=139201). .

[39] Created from NDB UD0017 (http://ndbserver.rutgers.edu/atlas/xray/structures/U/ud0017/ud0017.html)

[40] Greider C, Blackburn E (1985). "Identification of a specific telomere terminal transferase activity in Tetrahymena extracts". *Cell* **43** (2 Pt 1): 405–13. doi: 10.1016/0092-8674(85)90170-9 (http://dx.doi.org/10.1016/0092-8674(85)90170-9). PMID 3907856.

[41] Nugent C, Lundblad V (1998). " The telomerase reverse transcriptase: components and regulation (http://www.genesdev.org/cgi/content/full/12/8/1073)". *Genes Dev* **12** (8): 1073–85. doi: 10.1101/gad.12.8.1073 (http://dx.doi.org/10.1101/gad.12.8.1073). PMID 9553037. .

[42] Wright W, Tesmer V, Huffman K, Levene S, Shay J (1997). " Normal human chromosomes have long G-rich telomeric overhangs at one end (http://www.genesdev.org/cgi/content/full/11/21/2801)". *Genes Dev* **11** (21): 2801–9. doi: 10.1101/gad.11.21.2801 (http://dx.doi.org/10.1101/gad.11.21.2801). PMID 9353250. .

[43] Burge S, Parkinson G, Hazel P, Todd A, Neidle S (2006). " Quadruplex DNA: sequence, topology and structure (http://nar.oxfordjournals.org/cgi/pmidlookup?view=long&pmid=17012276)". *Nucleic Acids Res* **34** (19): 5402–15. doi: 10.1093/nar/gkl655 (http://dx.doi.org/10.1093/nar/gkl655). PMID 17012276. PMC: 1636468 (http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez&artid=1636468). .

[44] Parkinson G, Lee M, Neidle S (2002). "Crystal structure of parallel quadruplexes from human telomeric DNA". *Nature* **417** (6891): 876–80. doi: 10.1038/nature755 (http://dx.doi.org/10.1038/nature755). PMID 12050675.

[45] Griffith J, Comeau L, Rosenfield S, Stansel R, Bianchi A, Moss H, de Lange T (1999). "Mammalian telomeres end in a large duplex loop". *Cell* **97** (4): 503–14. doi: 10.1016/S0092-8674(00)80760-6 (http://dx.doi.org/10.1016/S0092-8674(00)80760-6). PMID 10338214.

[46] Seeman NC (November 2005). "DNA enables nanoscale control of the structure of matter". *Q. Rev. Biophys.* **38** (4): 363–71. doi: 10.1017/S0033583505004087 (http://dx.doi.org/10.1017/S0033583505004087). PMID 16515737.

[47] Klose R, Bird A (2006). "Genomic DNA methylation: the mark and its mediators". *Trends Biochem Sci* **31** (2): 89–97. doi: 10.1016/j.tibs.2005.12.008 (http://dx.doi.org/10.1016/j.tibs.2005.12.008). PMID 16403636.

[48] Bird A (2002). "DNA methylation patterns and epigenetic memory". *Genes Dev* **16** (1): 6–21. doi: 10.1101/gad.947102 (http://dx.doi.org/10.1101/gad.947102). PMID 11782440.

[49] Walsh C, Xu G (2006). "Cytosine methylation and DNA repair". *Curr Top Microbiol Immunol* **301**: 283–315. doi: 10.1007/3-540-31390-7_11 (http://dx.doi.org/10.1007/3-540-31390-7_11). PMID 16570853.

[50] Kriaucionis S, Heintz N (May 2009). "The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain". *Science* **324** (5929): 929–30. doi: 10.1126/science.1169786 (http://dx.doi.org/10.1126/science.1169786). PMID 19372393.

[51] Ratel D, Ravanat J, Berger F, Wion D (2006). "N6-methyladenine: the other methylated base of DNA". *Bioessays* **28** (3): 309–15. doi: 10.1002/bies.20342 (http://dx.doi.org/10.1002/bies.20342). PMID 16479578.

[52] Gommers-Ampt J, Van Leeuwen F, de Beer A, Vliegenthart J, Dizdaroglu M, Kowalak J, Crain P, Borst P (1993). "beta-D-glucosyl-hydroxymethyluracil: a novel modified base present in the DNA of the parasitic protozoan T. brucei". *Cell* **75** (6): 1129–36. doi: 10.1016/0092-8674(93)90322-H (http://dx.doi.org/10.1016/0092-8674(93)90322-H). PMID 8261512.
[53] Created from PDB 1JDG (http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1JDG)
[54] Douki T, Reynaud-Angelin A, Cadet J, Sage E (2003). "Bipyrimidine photoproducts rather than oxidative lesions are the main type of DNA damage involved in the genotoxic effect of solar UVA radiation". *Biochemistry* **42** (30): 9221–6. doi: 10.1021/bi034593c (http://dx.doi.org/10.1021/bi034593c). PMID 12885257.,
[55] Cadet J, Delatour T, Douki T, Gasparutto D, Pouget J, Ravanat J, Sauvaigo S (1999). "Hydroxyl radicals and DNA base damage". *Mutat Res* **424** (1–2): 9–21. PMID 10064846.
[56] Beckman KB, Ames BN (August 1997). " Oxidative decay of DNA (http://www.jbc.org/cgi/pmidlookup?view=long&pmid=9289489)". *J. Biol. Chem.* **272** (32): 19633–6. PMID 9289489. .
[57] Valerie K, Povirk L (2003). "Regulation and mechanisms of mammalian double-strand break repair". *Oncogene* **22** (37): 5792–812. doi: 10.1038/sj.onc.1206679 (http://dx.doi.org/10.1038/sj.onc.1206679). PMID 12947387.
[58] Ferguson L, Denny W (1991). "The genetic toxicology of acridines". *Mutat Res* **258** (2): 123–60. PMID 1881402.
[59] Jeffrey A (1985). "DNA modification by chemical carcinogens". *Pharmacol Ther* **28** (2): 237–72. doi: 10.1016/0163-7258(85)90013-0 (http://dx.doi.org/10.1016/0163-7258(85)90013-0). PMID 3936066.
[60] Stephens T, Bunde C, Fillmore B (2000). "Mechanism of action in thalidomide teratogenesis". *Biochem Pharmacol* **59** (12): 1489–99. doi: 10.1016/S0006-2952(99)00388-3 (http://dx.doi.org/10.1016/S0006-2952(99)00388-3). PMID 10799645.
[61] Braña M, Cacho M, Gradillas A, de Pascual-Teresa B, Ramos A (2001). "Intercalators as anticancer drugs". *Curr Pharm Des* **7** (17): 1745–80. doi: 10.2174/1381612013397113 (http://dx.doi.org/10.2174/1381612013397113). PMID 11562309.
[62] Venter J, *et al.* (2001). "The sequence of the human genome". *Science* **291** (5507): 1304–51. doi: 10.1126/science.1058040 (http://dx.doi.org/10.1126/science.1058040). PMID 11181995.
[63] Thanbichler M, Wang S, Shapiro L (2005). "The bacterial nucleoid: a highly organized and dynamic structure". *J Cell Biochem* **96** (3): 506–21. doi: 10.1002/jcb.20519 (http://dx.doi.org/10.1002/jcb.20519). PMID 15988757.
[64] Wolfsberg T, McEntyre J, Schuler G (2001). "Guide to the draft human genome". *Nature* **409** (6822): 824–6. doi: 10.1038/35057000 (http://dx.doi.org/10.1038/35057000). PMID 11236998.
[65] Gregory T (2005). " The C-value enigma in plants and animals: a review of parallels and an appeal for partnership (http://aob.oxfordjournals.org/cgi/content/full/95/1/133)". *Ann Bot (Lond)* **95** (1): 133–46. doi: 10.1093/aob/mci009 (http://dx.doi.org/10.1093/aob/mci009). PMID 15596463. .
[66] The ENCODE Project Consortium (2007). "Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project". *Nature* **447** (7146): 799–816. doi: 10.1038/nature05874 (http://dx.doi.org/10.1038/nature05874).
[67] Created from PDB 1MSW (http://www.rcsb.org/pdb/explore/explore.do?structureId=1MSW)
[68] Pidoux A, Allshire R (2005). " The role of heterochromatin in centromere function (http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez&artid=1569473)". *Philos Trans R Soc Lond B Biol Sci* **360** (1455): 569–79. doi: 10.1098/rstb.2004.1611 (http://dx.doi.org/10.1098/rstb.2004.1611). PMID 15905142.
[69] Harrison P, Hegyi H, Balasubramanian S, Luscombe N, Bertone P, Echols N, Johnson T, Gerstein M (2002). " Molecular fossils in the human genome: identification and analysis of the pseudogenes in chromosomes 21 and 22 (http://www.genome.org/cgi/content/full/12/2/272)". *Genome Res* **12** (2): 272–80. doi: 10.1101/gr.207102 (http://dx.doi.org/10.1101/gr.207102). PMID 11827946. .
[70] Harrison P, Gerstein M (2002). "Studying genomes through the aeons: protein families, pseudogenes and proteome evolution". *J Mol Biol* **318** (5): 1155–74. doi: 10.1016/S0022-2836(02)00109-2 (http://dx.doi.org/10.1016/S0022-2836(02)00109-2). PMID 12083509.
[71] Albà M (2001). " Replicative DNA polymerases (http://genomebiology.com/1465-6906/2/REVIEWS3002)". *Genome Biol* **2** (1): REVIEWS3002. doi: 10.1186/gb-2001-2-1-reviews3002 (http://dx.doi.org/10.1186/gb-2001-2-1-reviews3002). PMID 11178285. PMC: 150442 (http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez&artid=150442). .
[72] Sandman K, Pereira S, Reeve J (1998). "Diversity of prokaryotic chromosomal proteins and the origin of the nucleosome". *Cell Mol Life Sci* **54** (12): 1350–64. doi: 10.1007/s000180050259 (http://dx.doi.org/10.1007/s000180050259). PMID 9893710.
[73] Dame RT (2005). "The role of nucleoid-associated proteins in the organization and compaction of bacterial chromatin". *Mol. Microbiol.* **56** (4): 858–70. doi: 10.1111/j.1365-2958.2005.04598.x (http://dx.doi.org/10.

1111/j.1365-2958.2005.04598.x). PMID 15853876.
[74] Luger K, Mäder A, Richmond R, Sargent D, Richmond T (1997). "Crystal structure of the nucleosome core particle at 2.8 A resolution". *Nature* **389** (6648): 251–60. doi: 10.1038/38444 (http://dx.doi.org/10.1038/38444). PMID 9305837.
[75] Jenuwein T, Allis C (2001). "Translating the histone code". *Science* **293** (5532): 1074–80. doi: 10.1126/science.1063127 (http://dx.doi.org/10.1126/science.1063127). PMID 11498575.
[76] Ito T. "Nucleosome assembly and remodelling". *Curr Top Microbiol Immunol* **274**: 1–22. PMID 12596902.
[77] Thomas J (2001). "HMG1 and 2: architectural DNA-binding proteins". *Biochem Soc Trans* **29** (Pt 4): 395–401. doi: 10.1042/BST0290395 (http://dx.doi.org/10.1042/BST0290395). PMID 11497996.
[78] Grosschedl R, Giese K, Pagel J (1994). "HMG domain proteins: architectural elements in the assembly of nucleoprotein structures". *Trends Genet* **10** (3): 94–100. doi: 10.1016/0168-9525(94)90232-1 (http://dx.doi.org/10.1016/0168-9525(94)90232-1). PMID 8178371.
[79] Iftode C, Daniely Y, Borowiec J (1999). "Replication protein A (RPA): the eukaryotic SSB". *Crit Rev Biochem Mol Biol* **34** (3): 141–80. doi: 10.1080/10409239991209255 (http://dx.doi.org/10.1080/10409239991209255). PMID 10473346.
[80] Created from PDB 1LMB (http://www.rcsb.org/pdb/explore/explore.do?structureId=1LMB)
[81] Myers L, Kornberg R (2000). "Mediator of transcriptional regulation". *Annu Rev Biochem* **69**: 729–49. doi: 10.1146/annurev.biochem.69.1.729 (http://dx.doi.org/10.1146/annurev.biochem.69.1.729). PMID 10966474.
[82] Spiegelman B, Heinrich R (2004). "Biological control through regulated transcriptional coactivators". *Cell* **119** (2): 157–67. doi: 10.1016/j.cell.2004.09.037 (http://dx.doi.org/10.1016/j.cell.2004.09.037). PMID 15479634.
[83] Li Z, Van Calcar S, Qu C, Cavenee W, Zhang M, Ren B (2003). " A global transcriptional regulatory role for c-Myc in Burkitt's lymphoma cells (http://www.pnas.org/cgi/pmidlookup?view=long&pmid=12808131)". *Proc Natl Acad Sci USA* **100** (14): 8164–9. doi: 10.1073/pnas.1332764100 (http://dx.doi.org/10.1073/pnas.1332764100). PMID 12808131. PMC: 166200 (http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez&artid=166200). .
[84] Pabo C, Sauer R (1984). "Protein-DNA recognition". *Annu Rev Biochem* **53**: 293–321. doi: 10.1146/annurev.bi.53.070184.001453 (http://dx.doi.org/10.1146/annurev.bi.53.070184.001453). PMID 6236744.
[85] Created from PDB 1RVA (http://www.rcsb.org/pdb/explore/explore.do?structureId=1RVA)
[86] Bickle T, Krüger D (1993). " Biology of DNA restriction (http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez&artid=372918)". *Microbiol Rev* **57** (2): 434–50. PMID 8336674.
[87] Doherty A, Suh S (2000). " Structural and mechanistic conservation in DNA ligases (http://nar.oxfordjournals.org/cgi/pmidlookup?view=long&pmid=11058099)". *Nucleic Acids Res* **28** (21): 4051–8. doi: 10.1093/nar/28.21.4051 (http://dx.doi.org/10.1093/nar/28.21.4051). PMID 11058099. PMC: 113121 (http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez&artid=113121). .
[88] Schoeffler A, Berger J (2005). "Recent advances in understanding structure-function relationships in the type II topoisomerase mechanism". *Biochem Soc Trans* **33** (Pt 6): 1465–70. doi: 10.1042/BST20051465 (http://dx.doi.org/10.1042/BST20051465). PMID 16246147.
[89] Tuteja N, Tuteja R (2004). "Unraveling DNA helicases. Motif, structure, mechanism and function". *Eur J Biochem* **271** (10): 1849–63. doi: 10.1111/j.1432-1033.2004.04094.x (http://dx.doi.org/10.1111/j.1432-1033.2004.04094.x). PMID 15128295.
[90] Joyce C, Steitz T (1995). " Polymerase structures and function: variations on a theme? (http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez&artid=177480)". *J Bacteriol* **177** (22): 6321–9. PMID 7592405.
[91] Hubscher U, Maga G, Spadari S (2002). "Eukaryotic DNA polymerases". *Annu Rev Biochem* **71**: 133–63. doi: 10.1146/annurev.biochem.71.090501.150041 (http://dx.doi.org/10.1146/annurev.biochem.71.090501.150041). PMID 12045093.
[92] Johnson A, O'Donnell M (2005). "Cellular DNA replicases: components and dynamics at the replication fork". *Annu Rev Biochem* **74**: 283–315. doi: 10.1146/annurev.biochem.73.011303.073859 (http://dx.doi.org/10.1146/annurev.biochem.73.011303.073859). PMID 15952889.
[93] Tarrago-Litvak L, Andréola M, Nevinsky G, Sarih-Cottin L, Litvak S (01 May 1994). " The reverse transcriptase of HIV-1: from enzymology to therapeutic intervention (http://www.fasebj.org/cgi/reprint/8/8/497)". *Faseb J* **8** (8): 497–503. PMID 7514143. .
[94] Martinez E (2002). "Multi-protein complexes in eukaryotic gene transcription". *Plant Mol Biol* **50** (6): 925–47. doi: 10.1023/A:1021258713850 (http://dx.doi.org/10.1023/A:1021258713850). PMID 12516863.
[95] Created from PDB 1M6G (http://www.rcsb.org/pdb/explore/explore.do?structureId=1M6G)

[96] Cremer T, Cremer C (2001). "Chromosome territories, nuclear architecture and gene regulation in mammalian cells". *Nat Rev Genet* **2** (4): 292–301. doi: 10.1038/35066075 (http://dx.doi.org/10.1038/35066075). PMID 11283701.
[97] Pál C, Papp B, Lercher M (2006). "An integrated view of protein evolution". *Nat Rev Genet* **7** (5): 337–48. doi: 10.1038/nrg1838 (http://dx.doi.org/10.1038/nrg1838). PMID 16619049.
[98] O'Driscoll M, Jeggo P (2006). "The role of double-strand break repair - insights from human genetics". *Nat Rev Genet* **7** (1): 45–54. doi: 10.1038/nrg1746 (http://dx.doi.org/10.1038/nrg1746). PMID 16369571.
[99] Vispé S, Defais M (1997). "Mammalian Rad51 protein: a RecA homologue with pleiotropic functions". *Biochimie* **79** (9-10): 587–92. doi: 10.1016/S0300-9084(97)82007-X (http://dx.doi.org/10.1016/S0300-9084(97)82007-X). PMID 9466696.
[100] Neale MJ, Keeney S (2006). "Clarifying the mechanics of DNA strand exchange in meiotic recombination". *Nature* **442** (7099): 153–8. doi: 10.1038/nature04885 (http://dx.doi.org/10.1038/nature04885). PMID 16838012.
[101] Dickman M, Ingleston S, Sedelnikova S, Rafferty J, Lloyd R, Grasby J, Hornby D (2002). "The RuvABC resolvasome". *Eur J Biochem* **269** (22): 5492–501. doi: 10.1046/j.1432-1033.2002.03250.x (http://dx.doi.org/10.1046/j.1432-1033.2002.03250.x). PMID 12423347.
[102] Orgel L (2004). " Prebiotic chemistry and the origin of the RNA world (http://www.crbmb.com/cgi/reprint/39/2/99.pdf)" (PDF). *Crit Rev Biochem Mol Biol* **39** (2): 99–123. doi: 10.1080/10409230490460765 (http://dx.doi.org/10.1080/10409230490460765). PMID 15217990. .
[103] Davenport R (2001). "Ribozymes. Making copies in the RNA world". *Science* **292** (5520): 1278. doi: 10.1126/science.292.5520.1278a (http://dx.doi.org/10.1126/science.292.5520.1278a). PMID 11360970.
[104] Szathmáry E (1992). " What is the optimum size for the genetic alphabet? (http://www.pnas.org/cgi/reprint/89/7/2614.pdf)" (PDF). *Proc Natl Acad Sci USA* **89** (7): 2614–8. doi: 10.1073/pnas.89.7.2614 (http://dx.doi.org/10.1073/pnas.89.7.2614). PMID 1372984. .
[105] Lindahl T (1993). "Instability and decay of the primary structure of DNA". *Nature* **362** (6422): 709–15. doi: 10.1038/362709a0 (http://dx.doi.org/10.1038/362709a0). PMID 8469282.
[106] Vreeland R, Rosenzweig W, Powers D (2000). "Isolation of a 250 million-year-old halotolerant bacterium from a primary salt crystal". *Nature* **407** (6806): 897–900. doi: 10.1038/35038060 (http://dx.doi.org/10.1038/35038060). PMID 11057666.
[107] Hebsgaard M, Phillips M, Willerslev E (2005). "Geologically ancient DNA: fact or artefact?". *Trends Microbiol* **13** (5): 212–20. doi: 10.1016/j.tim.2005.03.010 (http://dx.doi.org/10.1016/j.tim.2005.03.010). PMID 15866038.
[108] Nickle D, Learn G, Rain M, Mullins J, Mittler J (2002). "Curiously modern DNA for a "250 million-year-old" bacterium". *J Mol Evol* **54** (1): 134–7. doi: 10.1007/s00239-001-0025-x (http://dx.doi.org/10.1007/s00239-001-0025-x). PMID 11734907.
[109] Goff SP, Berg P (1976). "Construction of hybrid viruses containing SV40 and lambda phage DNA segments and their propagation in cultured monkey cells". *Cell* **9** (4 PT 2): 695–705. doi: 10.1016/0092-8674(76)90133-1 (http://dx.doi.org/10.1016/0092-8674(76)90133-1). PMID 189942.
[110] Houdebine L. "Transgenic animal models in biomedical research". *Methods Mol Biol* **360**: 163–202. PMID 17172731.
[111] Daniell H, Dhingra A (2002). "Multigene engineering: dawn of an exciting new era in biotechnology". *Curr Opin Biotechnol* **13** (2): 136–41. doi: 10.1016/S0958-1669(02)00297-5 (http://dx.doi.org/10.1016/S0958-1669(02)00297-5). PMID 11950565.
[112] Job D (2002). "Plant biotechnology in agriculture". *Biochimie* **84** (11): 1105–10. doi: 10.1016/S0300-9084(02)00013-5 (http://dx.doi.org/10.1016/S0300-9084(02)00013-5). PMID 12595138.
[113] Collins A, Morton N (1994). " Likelihood ratios for DNA identification (http://www.pnas.org/cgi/reprint/91/13/6007.pdf)" (PDF). *Proc Natl Acad Sci USA* **91** (13): 6007–11. doi: 10.1073/pnas.91.13.6007 (http://dx.doi.org/10.1073/pnas.91.13.6007). PMID 8016106. .
[114] Weir B, Triggs C, Starling L, Stowell L, Walsh K, Buckleton J (1997). "Interpreting DNA mixtures". *J Forensic Sci* **42** (2): 213–22. PMID 9068179.
[115] Jeffreys A, Wilson V, Thein S (1985). "Individual-specific 'fingerprints' of human DNA". *Nature* **316** (6023): 76–9. doi: 10.1038/316076a0 (http://dx.doi.org/10.1038/316076a0). PMID 2989708.
[116] Colin Pitchfork — first murder conviction on DNA evidence also clears the prime suspect (http://www.forensic.gov.uk/forensic_t/inside/news/list_casefiles.php?case=1) Forensic Science Service Accessed 23 December 2006
[117] "DNA Identification in Mass Fatality Incidents" (http://massfatality.dna.gov/Introduction/). National Institute of Justice. September 2006. .
[118] Baldi, Pierre; Brunak, Soren (2001), *Bioinformatics: The Machine Learning Approach*, MIT Press, ISBN 978-0-262-02506-5, OCLC 45951728 57562233 (http://worldcat.org/oclc/45951728+57562233).

[119] Gusfield, Dan. *Algorithms on Strings, Trees, and Sequences: Computer Science and Computational Biology*. Cambridge University Press, 15 January 1997. ISBN 978-0-521-58519-4.
[120] Sjölander K (2004). " Phylogenomic inference of protein molecular function: advances and challenges (http://bioinformatics.oxfordjournals.org/cgi/reprint/20/2/170)". *Bioinformatics* **20** (2): 170–9. doi: 10.1093/bioinformatics/bth021 (http://dx.doi.org/10.1093/bioinformatics/bth021). PMID 14734307. .
[121] Mount DM (2004). *Bioinformatics: Sequence and Genome Analysis* (2 ed.). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press. ISBN 0879697121. OCLC 55106399 (http://worldcat.org/oclc/55106399).
[122] http://dx.doi.org/10.1371/journal.pbio.0020073
[123] Rothemund PW (March 2006). "Folding DNA to create nanoscale shapes and patterns". *Nature* **440** (7082): 297–302. doi: 10.1038/nature04586 (http://dx.doi.org/10.1038/nature04586). PMID 16541064.
[124] Andersen ES, Dong M, Nielsen MM, *et al.* (May 2009). "Self-assembly of a nanoscale DNA box with a controllable lid". *Nature* **459** (7243): 73–6. doi: 10.1038/nature07971 (http://dx.doi.org/10.1038/nature07971). PMID 19424153.
[125] Ishitsuka Y, Ha T (May 2009). "DNA nanotechnology: a nanomachine goes live". *Nat Nanotechnol* **4** (5): 281–2. doi: 10.1038/nnano.2009.101 (http://dx.doi.org/10.1038/nnano.2009.101). PMID 19421208.
[126] Aldaye FA, Palmer AL, Sleiman HF (September 2008). "Assembling materials with DNA as the guide". *Science* **321** (5897): 1795–9. doi: 10.1126/science.1154533 (http://dx.doi.org/10.1126/science.1154533). PMID 18818351.
[127] Wray G (2002). " Dating branches on the tree of life using DNA (http://genomebiology.com/1465-6906/3/REVIEWS0001)". *Genome Biol* **3** (1): REVIEWS0001. doi: 10.1046/j.1525-142X.1999.99010.x (http://dx.doi.org/10.1046/j.1525-142X.1999.99010.x). PMID 11806830. PMC: 150454 (http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez&artid=150454). .
[128] *Lost Tribes of Israel*, NOVA, PBS airdate: 22 February 2000. Transcript available from PBS.org, (http://www.pbs.org/wgbh/nova/transcripts/2706israel.html) (last accessed on 4 March 2006)
[129] Kleiman, Yaakov. "The Cohanim/DNA Connection: The fascinating story of how DNA studies confirm an ancient biblical tradition". (http://www.aish.com/societywork/sciencenature/the_cohanim_-_dna_connection.asp) *aish.com* (January 13, 2000). Accessed 4 March 2006.
[130] Bhattacharya, Shaoni. "Killer convicted thanks to relative's DNA". (http://www.newscientist.com/article.ns?id=dn4908) *newscientist.com* (20 April 2004). Accessed 22 December 06
[131] Dahm R (January 2008). "Discovering DNA: Friedrich Miescher and the early years of nucleic acid research". *Hum. Genet.* **122** (6): 565–81. doi: 10.1007/s00439-007-0433-0 (http://dx.doi.org/10.1007/s00439-007-0433-0). PMID 17901982.
[132] Levene P, (01 December 1919). " The structure of yeast nucleic acid (http://www.jbc.org/cgi/reprint/40/2/415)". *J Biol Chem* **40** (2): 415–24. .
[133] Astbury W, (1947). "Nucleic acid". *Symp. SOC. Exp. Bbl* **1** (66).
[134] Lorenz MG, Wackernagel W (01 September 1994). " Bacterial gene transfer by natural genetic transformation in the environment (http://mmbr.asm.org/cgi/pmidlookup?view=long&pmid=7968924)". *Microbiol. Rev.* **58** (3): 563–602. PMID 7968924. PMC: 372978 (http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez&artid=372978). .
[135] Avery O, MacLeod C, McCarty M (1944). " Studies on the chemical nature of the substance inducing transformation of pneumococcal types. Inductions of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III (http://www.jem.org/cgi/reprint/149/2/297)". *J Exp Med* **79** (2): 137–158. doi: 10.1084/jem.79.2.137 (http://dx.doi.org/10.1084/jem.79.2.137). .
[136] Hershey A, Chase M (1952). " Independent functions of viral protein and nucleic acid in growth of bacteriophage (http://www.jgp.org/cgi/reprint/36/1/39.pdf)" (PDF). *J Gen Physiol* **36** (1): 39–56. doi: 10.1085/jgp.36.1.39 (http://dx.doi.org/10.1085/jgp.36.1.39). PMID 12981234. .
[137] The B-DNA X-ray pattern on the right of this linked image (http://osulibrary.oregonstate.edu/specialcollections/coll/pauling/dna/pictures/sci9.001.5.html) was obtained by Rosalind Franklin and Raymond Gosling in May 1952 at high hydration levels of DNA and it has been labeled as "Photo 51"
[138] Nature Archives Double Helix of DNA: 50 Years (http://www.nature.com/nature/dna50/archive.html)
[139] Original X-ray diffraction image (http://osulibrary.oregonstate.edu/specialcollections/coll/pauling/dna/pictures/franklin-typeBphoto.html)
[140] The Nobel Prize in Physiology or Medicine 1962 (http://nobelprize.org/nobel_prizes/medicine/laureates/1962/) Nobelprize .org Accessed 22 December 06
[141] Brenda Maddox (23 January 2003). " The double helix and the 'wronged heroine' (http://www.biomath.nyu.edu/index/course/hw_articles/nature4.pdf)" (PDF). *Nature* **421**: 407–408. doi: 10.1038/nature01399 (http://dx.doi.org/10.1038/nature01399). PMID 12540909. .
[142] Crick, F.H.C. On degenerate templates and the adaptor hypothesis (PDF). (http://genome.wellcome.ac.uk/assets/wtx030893.pdf) genome.wellcome.ac.uk (Lecture, 1955). Accessed 22 December 2006

[143] Meselson M, Stahl F (1958). "The replication of DNA in *Escherichia coli*". *Proc Natl Acad Sci USA* **44** (7): 671–82. doi: 10.1073/pnas.44.7.671 (http://dx.doi.org/10.1073/pnas.44.7.671). PMID 16590258.
[144] The Nobel Prize in Physiology or Medicine 1968 (http://nobelprize.org/nobel_prizes/medicine/laureates/1968/) Nobelprize.org Accessed 22 December 06
[145] http://proteopedia.org/wiki/index.php/DNA

Further reading

- Calladine, Chris R.; Drew, Horace R.; Luisi, Ben F. and Travers, Andrew A. (2003). *Understanding DNA: the molecule & how it works*. Amsterdam: Elsevier Academic Press. ISBN 0-12-155089-3.
- Dennis, Carina; Julie Clayton (2003). *50 years of DNA*. Basingstoke: Palgrave Macmillan. ISBN 1-4039-1479-6.
- Judson, Horace Freeland (1996). *The eighth day of creation: makers of the revolution in biology*. Plainview, N.Y: CSHL Press. ISBN 0-87969-478-5.
- Olby, Robert C. (1994). *The path to the double helix: the discovery of DNA*. New York: Dover Publications. ISBN 0-486-68117-3., first published in October 1974 by MacMillan, with foreword by Francis Crick;the definitive DNA textbook,revised in 1994 with a 9 page postscript.
- Olby, Robert C. (2009). *Francis Crick: A Biography*. Plainview, N.Y: Cold Spring Harbor Laboratory Press. ISBN 0-87969-798-9.
- Ridley, Matt (2006). *Francis Crick: discoverer of the genetic code*. [Ashland, OH: Eminent Lives, Atlas Books. ISBN 0-06-082333-X.
- Berry, Andrew; Watson, James D. (2003). *DNA: the secret of life*. New York: Alfred A. Knopf. ISBN 0-375-41546-7.
- Stent, Gunther Siegmund; Watson, James D. (1980). *The double helix: a personal account of the discovery of the structure of DNA*. New York: Norton. ISBN 0-393-95075-1.
- Wilkins, Maurice (2003). *The third man of the double helix the autobiography of Maurice Wilkins*. Cambridge, Eng: University Press. ISBN 0-19-860665-6.

External links

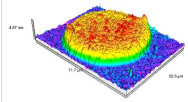
- DNA (http://www.dmoz.org/Science/Biology/Biochemistry_and_Molecular_Biology/Biomolecules/Nucleic_Acids/DNA//) at the Open Directory Project
- DNA binding site prediction on protein (http://pipe.scs.fsu.edu/displar.html)
- DNA coiling to form chromosomes (http://biostudio.com/c_ education mac.htm)
- DNA from the Beginning (http://www.dnaftb.org/dnaftb/) Another DNA Learning Center site on DNA, genes, and heredity from Mendel to the human genome project.
- DNA Lab, demonstrates how to extract DNA from wheat using readily available equipment and supplies. (http://ca.youtube.com/watch?v=iyb7fwduuGM)
- DNA the Double Helix Game (http://nobelprize.org/educational_games/medicine/dna_double_helix/) From the official Nobel Prize web site
- DNA under electron microscope (http://www.fidelitysystems.com/Unlinked_DNA.html)
- Dolan DNA Learning Center (http://www.dnalc.org/)
- Double Helix: 50 years of DNA (http://www.nature.com/nature/dna50/archive.html), *Nature*
- Double Helix 1953–2003 (http://www.ncbe.reading.ac.uk/DNA50/) National Centre for Biotechnology Education

- Francis Crick and James Watson talking on the BBC in 1962, 1972, and 1974 (http://www.bbc.co.uk/bbcfour/audiointerviews/profilepages/crickwatson1.shtml)
- Genetic Education Modules for Teachers (http://www.genome.gov/10506718) — *DNA from the Beginning* Study Guide
- Guide to DNA cloning (http://www.blackwellpublishing.com/trun/artwork/Animations/cloningexp/cloningexp.html)
- Olby R (January 2003). " Quiet debut for the double helix (http://chem-faculty.ucsd.edu/joseph/CHEM13/DNA1.pdf)". *Nature* **421** (6921): 402–5. doi: 10.1038/nature01397 (http://dx.doi.org/10.1038/nature01397). PMID 12540907. http://chem-faculty.ucsd.edu/joseph/CHEM13/DNA1.pdf.
- PDB Molecule of the Month *pdb23_1 (http://www.rcsb.org/pdb/static.do?p=education_discussion/molecule_of_the_month/pdb23_1.html)*
- Rosalind Franklin's contributions to the study of DNA (http://mason.gmu.edu/~emoody/rfranklin.html)
- The Register of Francis Crick Personal Papers 1938 - 2007 (http://orpheus.ucsd.edu/speccoll/testing/html/mss0660a.html#abstract) at Mandeville Special Collections Library, Geisel Library, University of California, San Diego
- U.S. National DNA Day (http://www.genome.gov/10506367) — watch videos and participate in real-time chat with top scientists
- " Clue to chemistry of heredity found (http://www.nytimes.com/packages/pdf/science/dna-article.pdf)". *The New York Times*. Saturday, June 13, 1953. http://www.nytimes.com/packages/pdf/science/dna-article.pdf. The first American newspaper coverage of the discovery of the DNA structure.

Molecular models of DNA

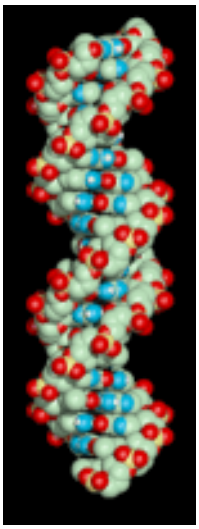
Molecular models of DNA structures are representations of the molecular geometry and topology of Deoxyribonucleic acid (→ DNA) molecules using one of several means, such as: closely packed spheres (CPK models) made of plastic, metal wires for 'skeletal models', graphic computations and animations by computers, artistic rendering, and so on, with the aim of simplifying and presenting the essential, physical and chemical, properties of DNA molecular structures either *in vivo* or *in vitro*. Computer molecular models also allow animations and molecular dynamics simulations that are very important for understanding how DNA functions *in vivo*. Thus, an old standing dynamic problem is how DNA "self-replication" takes place in living cells that should involve transient uncoiling of supercoiled DNA fibers. Although DNA consists of relatively rigid, very large elongated biopolymer molecules called "fibers" or chains (that are made of repeating nucleotide units of four basic types, attached to deoxyribose and phosphate groups), its molecular structure *in vivo* undergoes dynamic configuration changes that involve dynamically attached water molecules and ions. Supercoiling, packing with histones in chromosome structures, and other such supramolecular aspects also involve *in vivo* DNA topology which is even more complex than DNA molecular geometry, thus turning molecular modeling of DNA into an especially challenging problem for both molecular biologists and biotechnologists. Like other large molecules and biopolymers, DNA often exists in multiple stable geometries (that is, it exhibits conformational isomerism) and configurational, quantum states which are close to each other in energy on the potential energy surface of the DNA molecule. Such geometries can also be computed, at least in principle, by employing *ab initio* quantum chemistry methods that have high accuracy for small molecules. Such quantum geometries define an important class of *ab initio* molecular models of DNA whose exploration has barely started.

In an interesting twist of roles, the DNA molecule itself was proposed to be utilized for quantum computing. Both DNA nanostructures as well as DNA 'computing' biochips have been built (see biochip image at right).



DNA computing biochip:3D

The more advanced, computer-based molecular models of DNA involve molecular dynamics simulations as well as quantum mechanical computations of vibro-rotations, delocalized molecular orbitals (MOs), electric dipole moments, hydrogen-bonding, and so on.



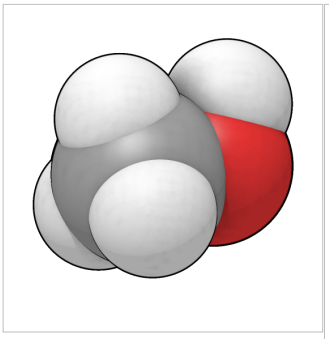
Spinning DNA generic model.

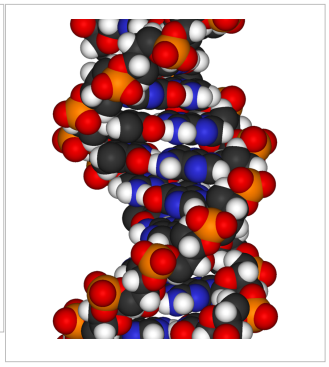
Importance

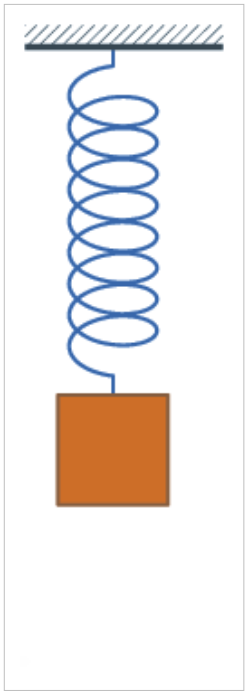
From the very early stages of structural studies of DNA by X-ray diffraction and biochemical means, molecular models such as the Watson-Crick double-helix model were successfully employed to solve the 'puzzle' of DNA structure, and also find how the latter relates to its key functions in living cells. The first high quality X-ray diffraction patterns of A-DNA were reported by Rosalind Franklin and Raymond Gosling in 1953[1] . The first calculations of the Fourier transform of an atomic helix were reported one year earlier by Cochran, Crick and Vand [2] , and were followed in 1953 by the computation of the Fourier transform of a coiled-coil by Crick[3] . The first reports of a double-helix molecular model of B-DNA structure were made by Watson and Crick in 1953[4] [5] . Last-but-not-least, Maurice F. Wilkins, A. Stokes and H.R. Wilson, reported the first X-ray patterns of *in vivo* B-DNA in partially oriented salmon sperm heads [30] . The development of the first correct double-helix molecular model of DNA by Crick and Watson may not have been possible without the biochemical evidence for the nucleotide base-pairing ([A---T]; [C---G]), or Chargaff's rules[6] [7] [8] [9] [10] [11] .

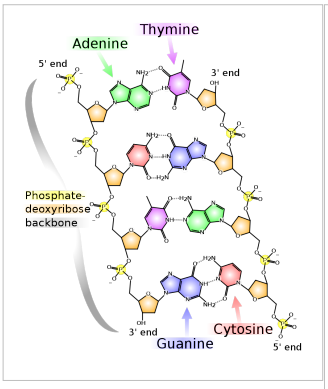
Examples of DNA molecular models

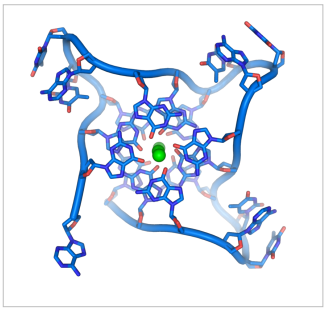
Animated molecular models allow one to visually explore the three-dimensional (3D) structure of DNA. The first DNA model is a space-filling, or CPK, model of the DNA double-helix whereas the third is an animated wire, or skeletal type, molecular model of DNA. The last two DNA molecular models in this series depict quadruplex DNA [12] that may be involved in certain cancers[13] [14] . The last figure on this panel is a molecular model of hydrogen bonds between water molecules in ice that are similar to those found in DNA.

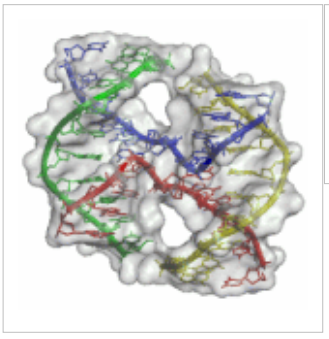


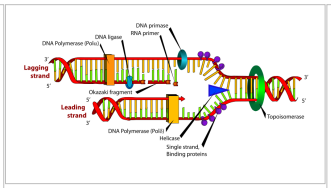


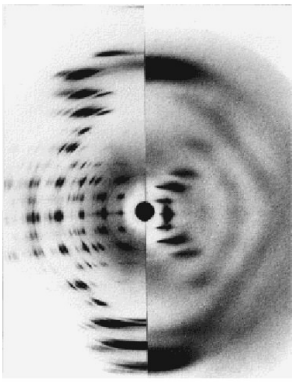



Thymine
Adenine
5' end
3' end
Phosphate-
deoxyribose
backbone
3' end
Guanine
Cytosine
5' end

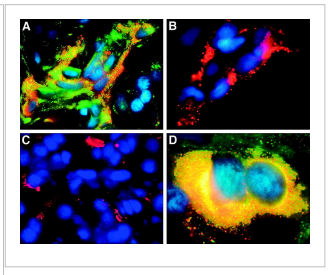


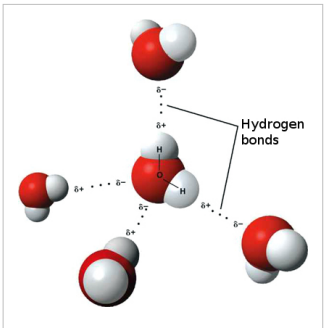



Lagging strand
Leading strand
Okazaki fragment
DNA Polymerase (Pol I)
DNA ligase
DNA primase
RNA primer
DNA Polymerase (Pol III)
Helicase
Single strand. Binding proteins
Topoisomerase

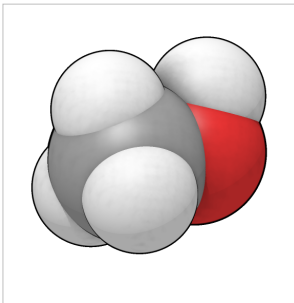


A-DNA B-DNA


A
B
C
D



- Spacefilling model or CPK model - a molecule is represented by overlapping spheres representing the atoms.

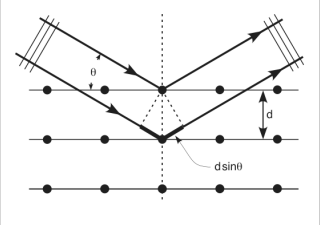


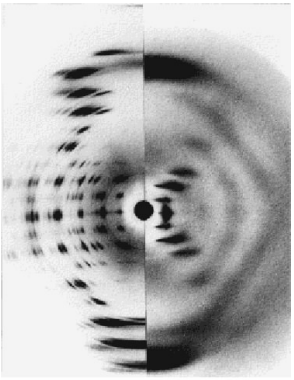
Images for DNA Structure Determination from X-Ray Patterns

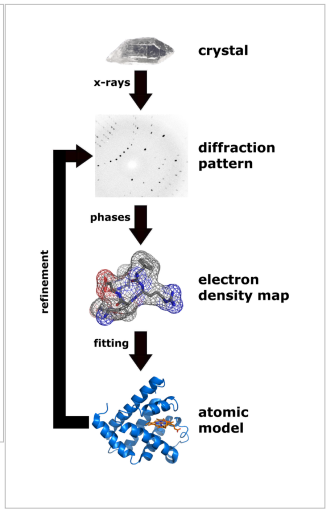
The following images illustrate both the principles and the main steps involved in generating structural information from X-ray diffraction studies of oriented DNA fibers with the help of molecular models of DNA that are combined with crystallographic and mathematical analysis of the X-ray patterns. From left to right the gallery of images shows:

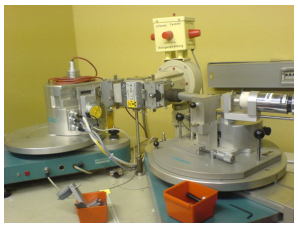
- *First row*:
- 1. Constructive X-ray interference, or diffraction, following Bragg's Law of X-ray "*reflection by the crystal planes*";
- 2. A comparison of A-DNA (crystalline) and highly hydrated B-DNA (paracrystalline) X-ray diffraction, and respectively, X-ray scattering patterns (courtesy of Dr. Herbert R. Wilson, FRS- see refs. list);
- 3. Purified DNA precipitated in a water jug;
- 4. The major steps involved in DNA structure determination by X-ray crystallography showing the important role played by molecular models of DNA structure in this iterative, structure--determination process;
- *Second row*:
- 5. Photo of a modern X-ray diffractometer employed for recording X-ray patterns of DNA with major components: X-ray source, goniometer, sample holder, X-ray detector and/or plate holder;
- 6. Illustrated animation of an X-ray goniometer;
- 7. X-ray detector at the SLAC synchrotron facility;
- 8. Neutron scattering facility at ISIS in UK;
- *Third and fourth rows*: *Molecular models of DNA structure at various scales*; figure #11 is an actual electron micrograph of a DNA fiber bundle, presumably of a single

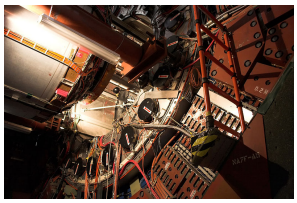
bacterial chromosome loop.

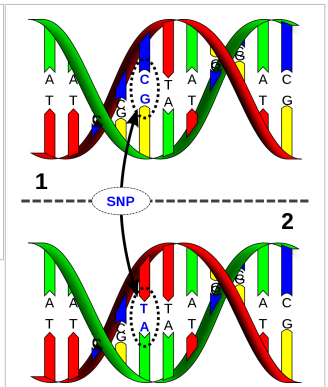


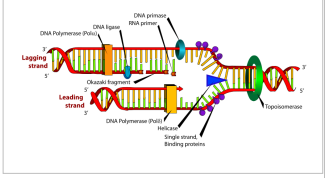


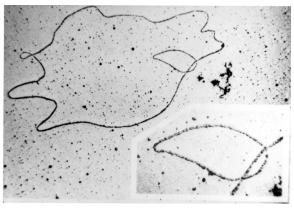


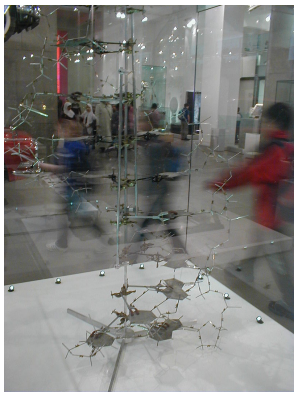


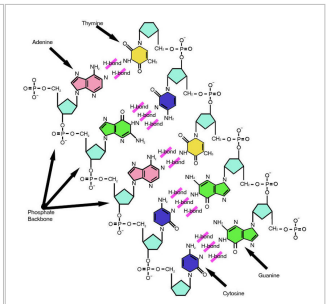


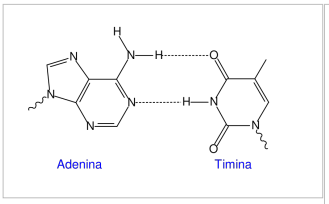


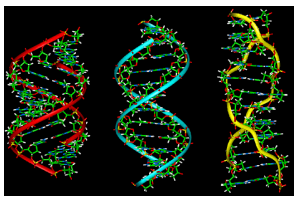


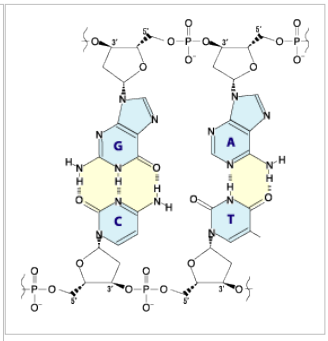


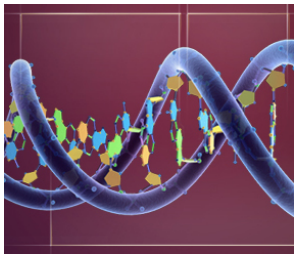


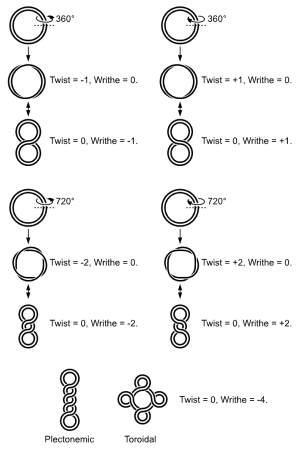


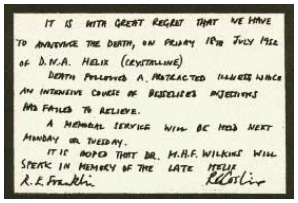






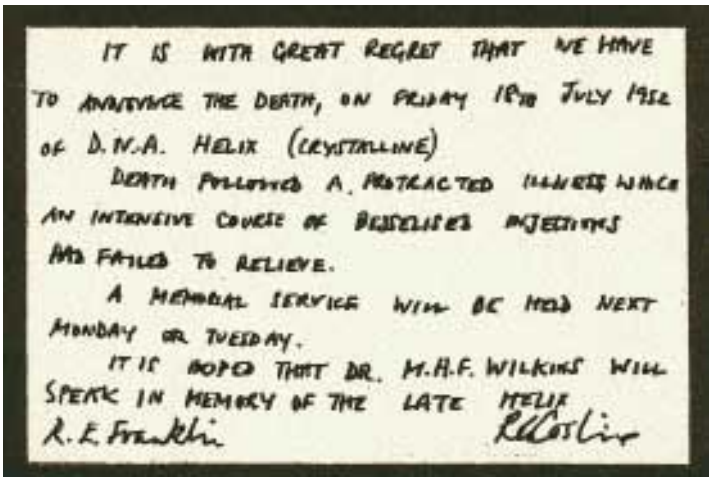






Paracrystalline lattice models of B-DNA structures

A paracrystalline lattice, or paracrystal, is a molecular or atomic lattice with significant amounts (e.g., larger than a few percent) of partial disordering of molecular arranegements. Limiting cases of the paracrystal model are nanostructures, such as glasses, liquids, etc., that may possess only local ordering and no global order. Liquid crystals also have paracrystalline rather than crystalline structures.



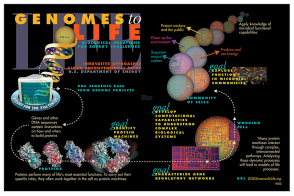
DNA Helix controversy in 1952

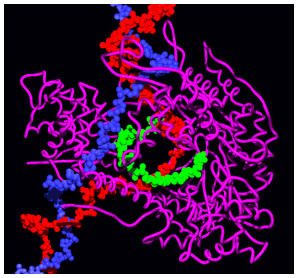
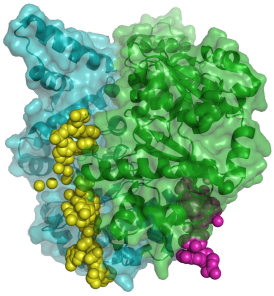
Highly hydrated B-DNA occurs naturally in living cells in such a paracrystalline state, which is a dynamic one in spite of the relatively rigid DNA double-helix stabilized by parallel hydrogen bonds between the nucleotide base-pairs in the two complementary, helical DNA chains (see figures). For simplicity most DNA molecular models ommit both water and ions dynamically bound to B-DNA, and are thus less useful for understanding the dynamic behaviors of B-DNA *in vivo*. The physical and mathematical analysis of X-ray[15] [16] and spectroscopic data for paracrystalline B-DNA is therefore much more complicated than that of crystalline, A-DNA X-ray diffraction patterns. The paracrystal model is also important for DNA technological applications such as DNA nanotechnology. Novel techniques that combine X-ray diffraction of DNA with X-ray microscopy in hydrated living cells are now also being developed (see, for example, "Application of X-ray microscopy in the analysis of living hydrated cells" [17]).

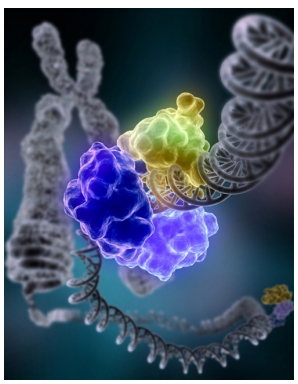
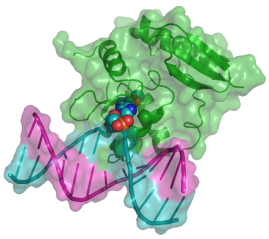
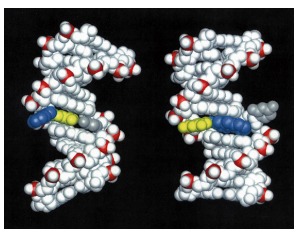
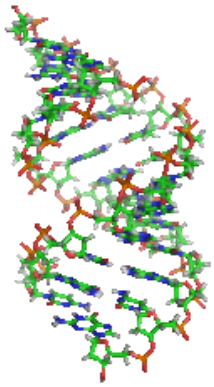
Genomic and Biotechnology Applications of DNA molecular modeling

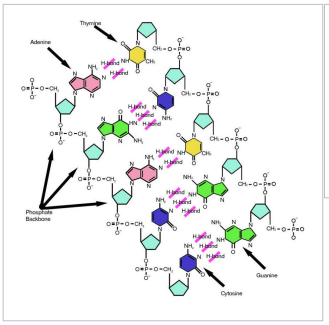
The following gallery of images illustrates various uses of DNA molecular modeling in Genomics and Biotechnology research applications from DNA repair to PCR and DNA nanostructures; each slide contains its own explanation and/or details. The first slide presents an overview of DNA applications, including DNA molecular models, with emphasis on Genomics and Biotechnology.

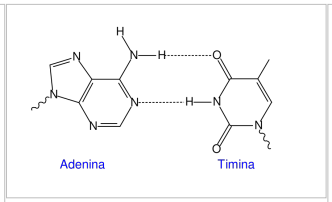
Gallery: DNA Molecular modeling applications

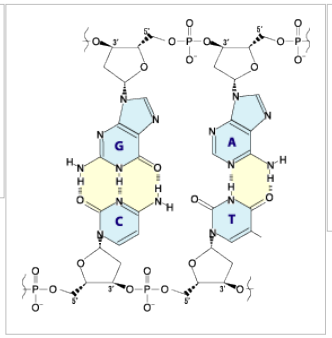


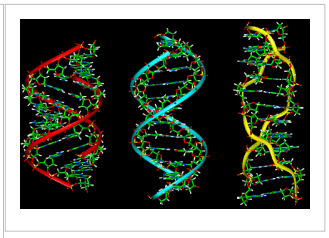



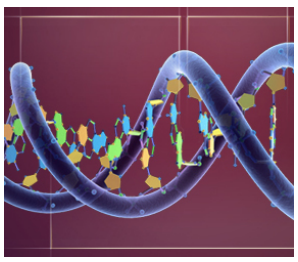





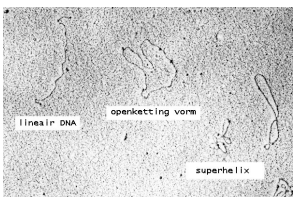

Thymine
Adenine
Phosphate Backbone
Guanine
Cytosine

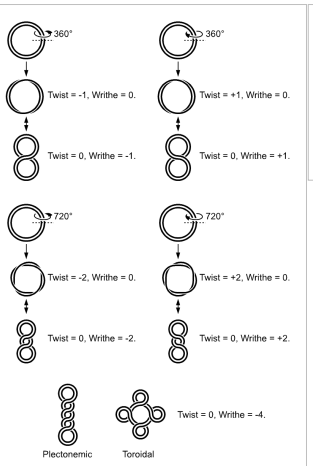

Adenina
Timina

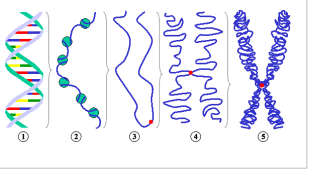

G
C
A
T

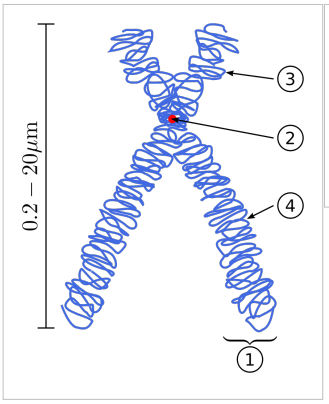


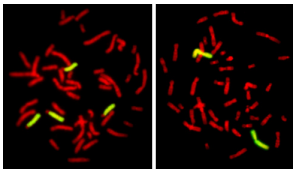


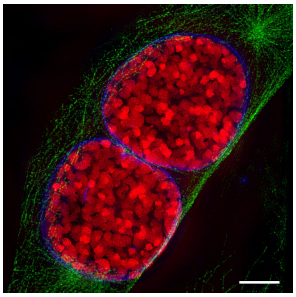

linear DNA
openkettling vorm
superhelix

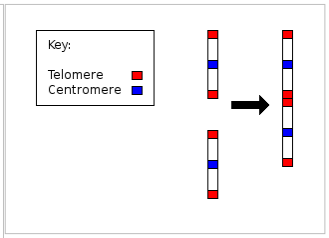

360°
Twist = -1, Writhe = 0.
Twist = 0, Writhe = -1.
360°
Twist = +1, Writhe = 0.
Twist = 0, Writhe = +1.
720°
Twist = -2, Writhe = 0.
Twist = 0, Writhe = -2.
720°
Twist = +2, Writhe = 0.
Twist = 0, Writhe = +2.
Twist = 0, Writhe = -4.
Plectonemic
Toroidal

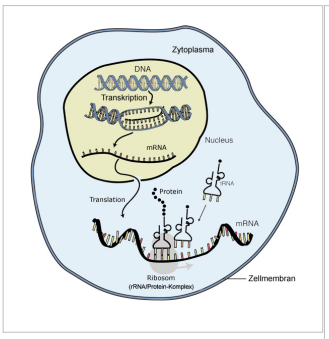

1
2
3
4
5

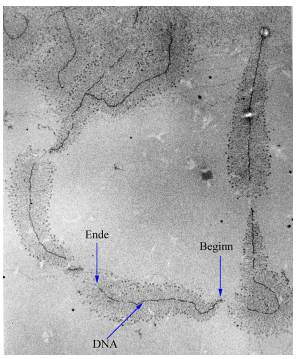

0.2 – 20μm
1
2
3
4

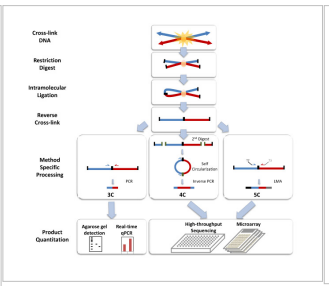


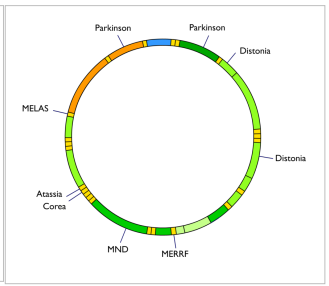


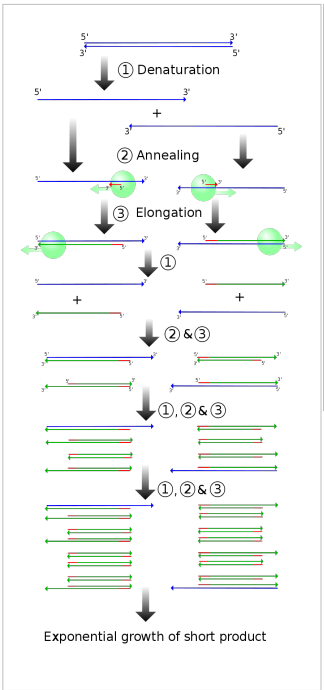

Key:
Telomere
Centromere

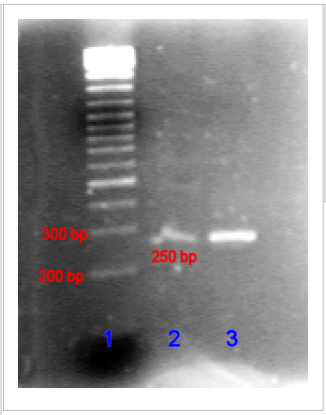

Zytoplasma
DNA
Transkription
Nucleus
mRNA
Translation
Protein
tRNA
mRNA
Ribosom
(rRNA/Protein-Komplex)
Zellmembran

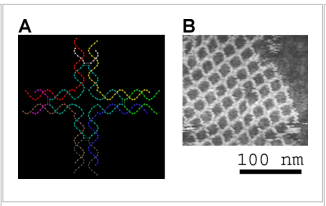

Ende
Beginn
DNA

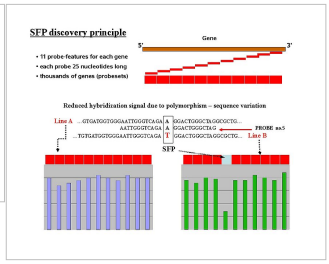

Cross-link DNA
Restriction Digest
Intramolecular Ligation
Reverse Cross-link
Method Specific Processing
3C
4C
5C
Product Quantitation
Agarose gel detection
Real-time qPCR
High-throughput Sequencing
Microarray

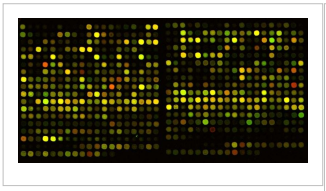

Parkinson
Parkinson
Distonia
MELAS
Distonia
Atassia
Corea
MND
MERRF

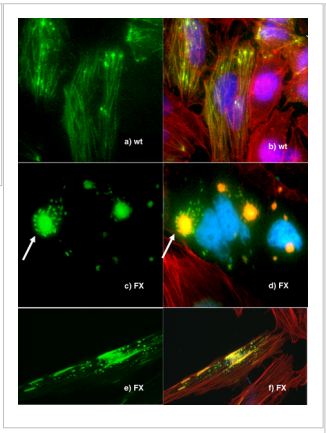


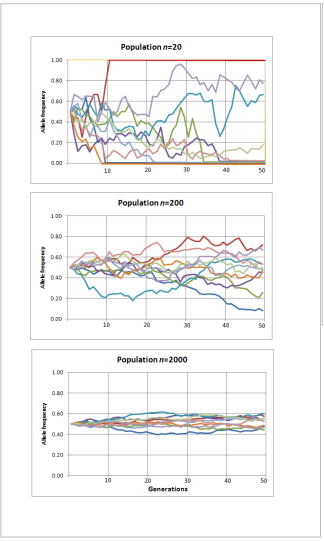












Databases for DNA molecular models and sequences

X-ray diffraction

- NDB ID: UD0017 Database [18]
- X-ray Atlas -database [19]
- PDB files of coordinates for nucleic acid structures from X-ray diffraction by NA (incl. DNA) crystals [20]
- Structure factors dowloadable files in CIF format [21]

Neutron scattering

- ISIS neutron source
- ISIS pulsed neutron source:A world centre for science with neutrons & muons at Harwell, near Oxford, UK. [22]

X-ray microscopy

- Application of X-ray microscopy in the analysis of living hydrated cells [23]

Electron microscopy

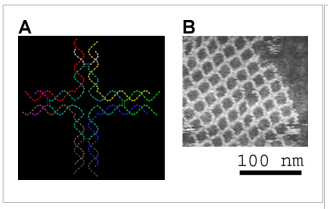
- DNA under electron microscope [24]

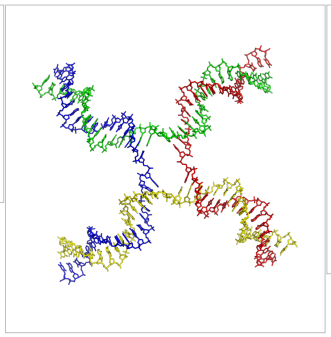
Atomic Force Microscopy (AFM)

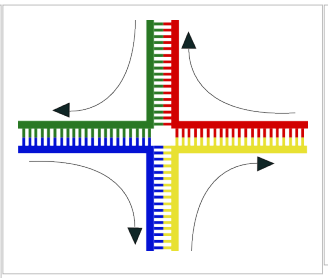
Two-dimensional DNA junction arrays have been visualized by Atomic Force Microscopy (AFM)[25] . Other imaging resources for AFM/Scanning probe microscopy(SPM) can be freely accessed at:

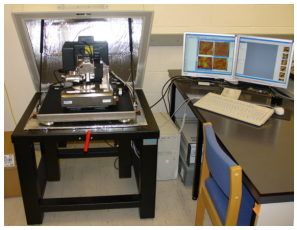
- How SPM Works [26]
- SPM Image Gallery - AFM STM SEM MFM NSOM and more. [27]

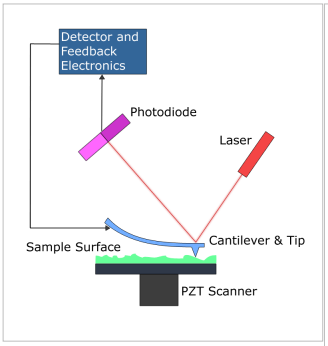
Gallery of AFM Images

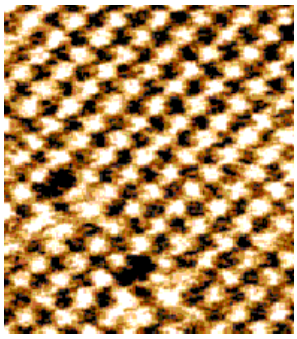


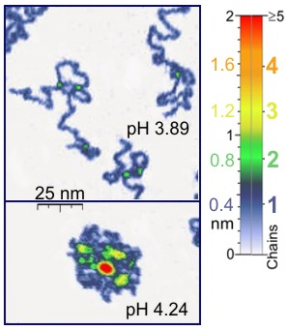


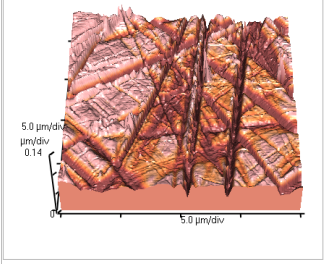




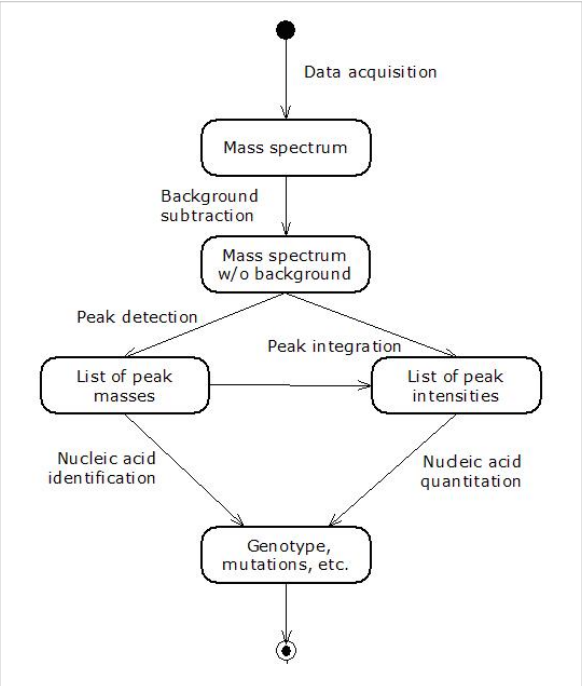








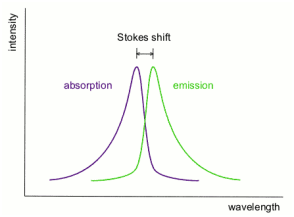
Mass spectrometry--Maldi informatics

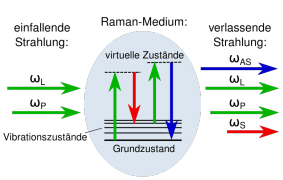


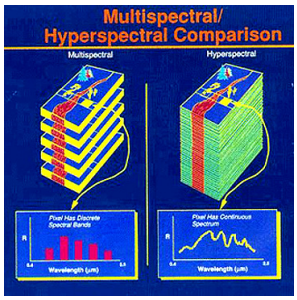
Spectroscopy

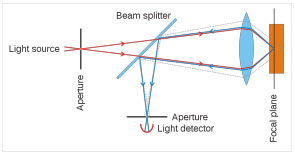
- Vibrational circular dichroism (VCD)
- FT-NMR[28] [29]
  - NMR Atlas--database [30]
  - mmcif downloadable coordinate files of nucleic acids in solution from 2D-FT NMR data [31]
  - NMR constraints files for NAs in PDB format [32]
- NMR microscopy[33]
- Microwave spectroscopy
- FT-IR
- FT-NIR[34] [35] [36]
- Spectral, Hyperspectral, and Chemical imaging)[37] [38] [39] [40] [41] [42] [43] .
- Raman spectroscopy/microscopy[44] and CARS[45] .
- Fluorescence correlation spectroscopy[46] [47] [48] [49] [50] [51] [52] [53] , Fluorescence cross-correlation spectroscopy and FRET[54] [55] [56] .
- Confocal microscopy[57]

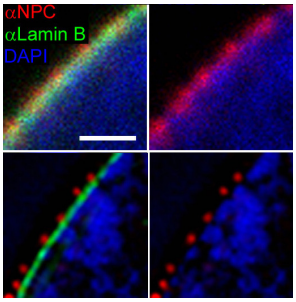
Gallery: CARS (Raman spectroscopy), Fluorescence confocal microscopy, and Hyperspectral imaging

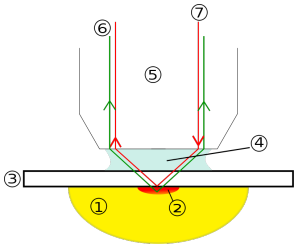


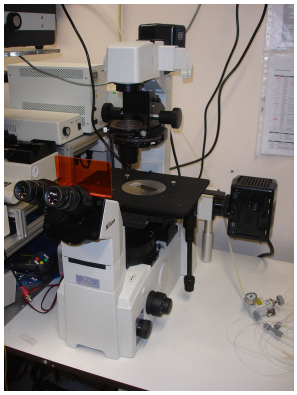


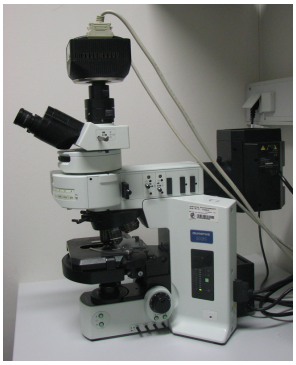


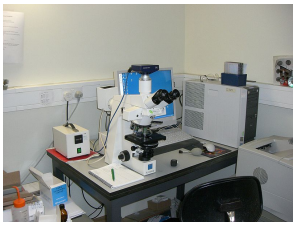


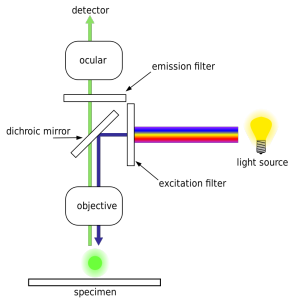


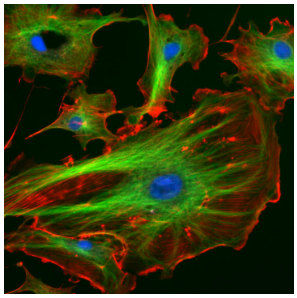


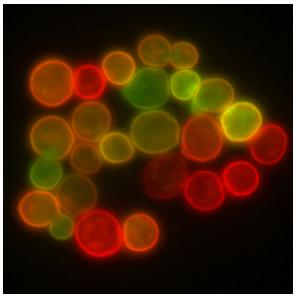


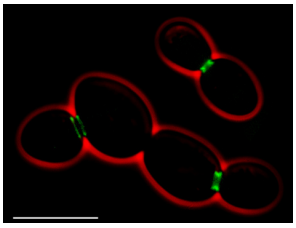


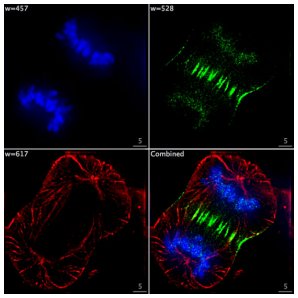


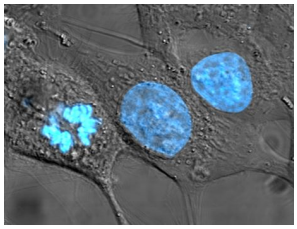


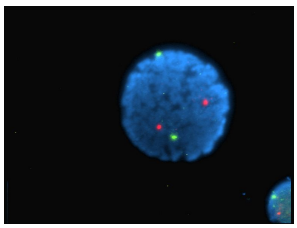
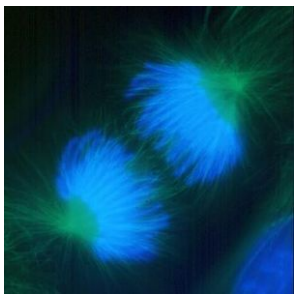
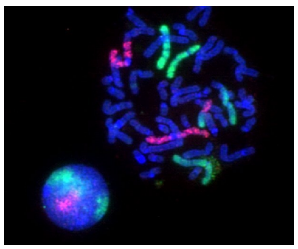
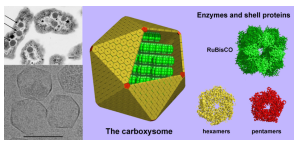










Genomic and structural databases

- CBS Genome Atlas Database [58] — contains examples of base skews.[59]
- The Z curve database of genomes — a 3-dimensional visualization and analysis tool of genomes [60][61] .
- DNA and other nucleic acids' molecular models: Coordinate files of nucleic acids molecular structure models in PDB and CIF formats [62]

Notes

[1] Franklin, R.E. and Gosling, R.G. recd.6 March 1953. Acta Cryst. (1953). 6, 673 The Structure of Sodium Thymonucleate Fibres I. The Influence of Water Content Acta Cryst. (1953). and 6, 678 The Structure of Sodium Thymonucleate Fibres II. The Cylindrically Symmetrical Patterson Function.
[2] Cochran, W., Crick, F.H.C. and Vand V. 1952. The Structure of Synthetic Polypeptides. 1. The Transform of Atoms on a Helix. Acta Cryst. 5(5):581-586.
[3] Crick, F.H.C. 1953a. The Fourier Transform of a Coiled-Coil., Acta Crystallographica 6(8-9):685-689.
[4] Watson, J.D; Crick F.H.C. 1953a. Molecular Structure of Nucleic Acids- A Structure for Deoxyribose Nucleic Acid., Nature 171(4356):737-738.
[5] Watson, J.D; Crick F.H.C. 1953b. The Structure of DNA., Cold Spring Harbor Symposia on Qunatitative Biology 18:123-131.
[6] Elson D, Chargaff E (1952). "On the deoxyribonucleic acid content of sea urchin gametes". *Experientia* **8** (4): 143-145.
[7] Chargaff E, Lipshitz R, Green C (1952). "Composition of the deoxypentose nucleic acids of four genera of sea-urchin". *J Biol Chem* **195** (1): 155-160. PMID 14938364.
[8] Chargaff E, Lipshitz R, Green C, Hodes ME (1951). "The composition of the deoxyribonucleic acid of salmon sperm". *J Biol Chem* **192** (1): 223-230. PMID 14917668.
[9] Chargaff E (1951). "Some recent studies on the composition and structure of nucleic acids". *J Cell Physiol Suppl* **38** (Suppl).
[10] Magasanik B, Vischer E, Doniger R, Elson D, Chargaff E (1950). "The separation and estimation of ribonucleotides in minute quantities". *J Biol Chem* **186** (1): 37-50. PMID 14778802.
[11] Chargaff E (1950). "Chemical specificity of nucleic acids and mechanism of their enzymatic degradation". *Experientia* **6** (6): 201-209.
[12] http://ndbserver.rutgers.edu/atlas/xray/structures/U/ud0017/ud0017.html
[13] http://www.phy.cam.ac.uk/research/bss/molbiophysics.php
[14] http://planetphysics.org/encyclopedia/TheoreticalBiophysics.html
[15] Hosemann R., Bagchi R.N., *Direct analysis of diffraction by matter*, North-Holland Publs., Amsterdam – New York, 1962.
[16] Baianu, I.C. (1978). "X-ray scattering by partially disordered membrane systems.". *Acta Cryst.,* **A34** (5): 751–753. doi: 10.1107/S0567739478001540 (http://dx.doi.org/10.1107/S0567739478001540).
[17] http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=12379938
[18] http://ndbserver.rutgers.edu/atlas/xray/structures/U/ud0017/ud0017.html
[19] http://ndbserver.rutgers.edu/atlas/xray/index.html
[20] http://ndbserver.rutgers.edu/ftp/NDB/coordinates/na-biol/
[21] http://ndbserver.rutgers.edu/ftp/NDB/structure-factors/

[22] http://www.isis.rl.ac.uk/
[23] http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=12379938
[24] http://www.fidelitysystems.com/Unlinked_DNA.html
[25] Mao, Chengde; Sun, Weiqiong & Seeman, Nadrian C. (16 June 1999). "Designed Two-Dimensional DNA Holliday Junction Arrays Visualized by Atomic Force Microscopy". *Journal of the American Chemical Society* **121** (23): 5437–5443. doi: 10.1021/ja9900398 (http://dx.doi.org/10.1021/ja9900398). ISSN 0002-7863 (http://worldcat.org/issn/0002-7863).
[26] http://www.parkafm.com/New_html/resources/01general.php
[27] http://www.rhk-tech.com/results/showcase.php
[28] (http://www.jonathanpmiller.com/Karplus.html)- obtaining dihedral angles from ^{3}J coupling constants
[29] (http://www.spectroscopynow.com/FCKeditor/UserFiles/File/specNOW/HTML files/General_Karplus_Calculator.htm) Another Javascript-like NMR coupling constant to dihedral
[30] http://ndbserver.rutgers.edu/atlas/nmr/index.html
[31] http://ndbserver.rutgers.edu/ftp/NDB/coordinates/na-nmr-mmcif/
[32] http://ndbserver.rutgers.edu/ftp/NDB/nmr-restraints/
[33] Lee, S. C. et al., (2001). One Micrometer Resolution NMR Microscopy. *J. Magn. Res.*, **150**: 207-213.
[34] Near Infrared Microspectroscopy, Fluorescence Microspectroscopy,Infrared Chemical Imaging and High Resolution Nuclear Magnetic Resonance Analysis of Soybean Seeds, Somatic Embryos and Single Cells., Baianu, I.C. et al. 2004., In *Oil Extraction and Analysis.*, D. Luthria, Editor pp.241-273, AOCS Press., Champaign, IL.
[35] Single Cancer Cell Detection by Near Infrared Microspectroscopy, Infrared Chemical Imaging and Fluorescence Microspectroscopy.2004.I. C. Baianu, D. Costescu, N. E. Hofmann and S. S. Korban, q-bio/0407006 (July 2004) (http://arxiv.org/abs/q-bio/0407006)
[36] Raghavachari, R., Editor. 2001. *Near-Infrared Applications in Biotechnology*, Marcel-Dekker, New York, NY.
[37] http://www.imaging.net/chemical-imaging/ Chemical imaging
[38] http://www.malvern.com/LabEng/products/sdi/bibliography/sdi_bibliography.htm E. N. Lewis, E. Lee and L. H. Kidder, Combining Imaging and Spectroscopy: Solving Problems with Near-Infrared Chemical Imaging. Microscopy Today, Volume 12, No. 6, 11/2004.
[39] D.S. Mantus and G. H. Morrison. 1991. Chemical imaging in biology and medicine using ion microscopy., *Microchimica Acta*, **104**, (1-6) January 1991, doi: 10.1007/BF01245536
[40] Near Infrared Microspectroscopy, Fluorescence Microspectroscopy,Infrared Chemical Imaging and High Resolution Nuclear Magnetic Resonance Analysis of Soybean Seeds, Somatic Embryos and Single Cells., Baianu, I.C. et al. 2004., In *Oil Extraction and Analysis.*, D. Luthria, Editor pp.241-273, AOCS Press., Champaign, IL.
[41] Single Cancer Cell Detection by Near Infrared Microspectroscopy, Infrared Chemical Imaging and Fluorescence Microspectroscopy.2004.I. C. Baianu, D. Costescu, N. E. Hofmann and S. S. Korban, q-bio/0407006 (July 2004) (http://arxiv.org/abs/q-bio/0407006)
[42] J. Dubois, G. Sando, E. N. Lewis, Near-Infrared Chemical Imaging, A Valuable Tool for the Pharmaceutical Industry, G.I.T. Laboratory Journal Europe, No.1-2, 2007.
[43] Applications of Novel Techniques to Health Foods, Medical and Agricultural Biotechnology.(June 2004).,I. C. Baianu, P. R. Lozano, V. I. Prisecaru and H. C. Lin q-bio/0406047 (http://arxiv.org/abs/q-bio/0406047)
[44] Chemical Imaging Without Dyeing (http://witec.de/en/download/Raman/ImagingMicroscopy04.pdf)
[45] C.L. Evans and X.S. Xie.2008. Coherent Anti-Stokes Raman Scattering Microscopy: Chemical Imaging for Biology and Medicine., doi:10.1146/annurev.anchem.1.031207.112754 *Annual Review of Analytical Chemistry*, **1**: 883-909.
[46] Eigen, M., Rigler, M. Sorting single molecules: application to diagnostics and evolutionary biotechnology,(1994) *Proc. Natl. Acad. Sci. USA*, 91,5740-5747.
[47] Rigler, M. Fluorescence correlations, single molecule detection and large number screening. Applications in biotechnology,(1995) J. Biotechnol., 41,177-186.
[48] Rigler R. and Widengren J. (1990). Ultrasensitive detection of single molecules by fluorescence correlation spectroscopy, *BioScience* (Ed. Klinge & Owman) p.180.
[49] Single Cancer Cell Detection by Near Infrared Microspectroscopy, Infrared Chemical Imaging and Fluorescence Microspectroscopy.2004.I. C. Baianu, D. Costescu, N. E. Hofmann, S. S. Korban and et al., q-bio/0407006 (July 2004) (http://arxiv.org/abs/q-bio/0407006)
[50] Oehlenschläger F., Schwille P. and Eigen M. (1996). Detection of HIV-1 RNA by nucleic acid sequence-based amplification combined with fluorescence correlation spectroscopy, *Proc. Natl. Acad. Sci. USA* **93**:1281.
[51] Bagatolli, L.A., and Gratton, E. (2000). Two-photon fluorescence microscopy of coexisting lipid domains in giant unilamellar vesicles of binary phospholipid mixtures. Biophys J., 78:290-305.

[52] Schwille, P., Haupts, U., Maiti, S., and Webb. W.(1999). Molecular dynamics in living cells observed by fluorescence correlation spectroscopy with one- and two-photon excitation. Biophysical Journal, 77(10):2251-2265.
[53] Near Infrared Microspectroscopy, Fluorescence Microspectroscopy,Infrared Chemical Imaging and High Resolution Nuclear Magnetic Resonance Analysis of Soybean Seeds, Somatic Embryos and Single Cells., Baianu, I.C. et al. 2004., In *Oil Extraction and Analysis.*, D. Luthria, Editor pp.241-273, AOCS Press., Champaign, IL.
[54] FRET description (http://dwb.unl.edu/Teacher/NSF/C08/C08Links/pps99.cryst.bbk.ac.uk/projects/gmocz/fret.htm)
[55] doi:10.1016/S0959-440X(00)00190-1 (http://dx.doi.org/10.1016/S0959-440X(00)00190-1)Recent advances in FRET: distance determination in protein–DNA complexes. *Current Opinion in Structural Biology* **2001**, *11(2)*, 201-207
[56] http://www.fretimaging.org/mcnamaraintro.html FRET imaging introduction
[57] Eigen, M., and Rigler, R. (1994). Sorting single molecules: Applications to diagnostics and evolutionary biotechnology, Proc. Natl. Acad. Sci. USA 91:5740.
[58] http://www.cbs.dtu.dk/services/GenomeAtlas/
[59] Hallin PF, David Ussery D (2004). "CBS Genome Atlas Database: A dynamic storage for bioinformatic results and DNA sequence data". *Bioinformatics* **20**: 3682-3686.
[60] http://tubic.tju.edu.cn/zcurve/
[61] Zhang CT, Zhang R, Ou HY (2003). "The Z curve database: a graphic representation of genome sequences". Bioinformatics 19 (5): 593-599. doi:10.1093/bioinformatics/btg041
[62] http://ndbserver.rutgers.edu/ftp/NDB/models/

References

- *Applications of Novel Techniques to Health Foods, Medical and Agricultural Biotechnology*.(June 2004) I. C. Baianu, P. R. Lozano, V. I. Prisecaru and H. C. Lin., q-bio/0406047.
- F. Bessel, *Untersuchung des Theils der planetarischen Störungen*, Berlin Abhandlungen (1824), article 14.
- Sir Lawrence Bragg, FRS. *The Crystalline State, A General survey.* London: G. Bells and Sons, Ltd., vols. 1 and 2., 1966., 2024 pages.
- Cantor, C. R. and Schimmel, P.R. *Biophysical Chemistry, Parts I and II.*, San Franscisco: W.H. Freeman and Co. 1980. 1,800 pages.
- Eigen, M., and Rigler, R. (1994). Sorting single molecules: Applications to diagnostics and evolutionary biotechnology, *Proc. Natl. Acad. Sci. USA* **91**:5740.
- Raghavachari, R., Editor. 2001. *Near-Infrared Applications in Biotechnology*, Marcel-Dekker, New York, NY.
- Rigler R. and Widengren J. (1990). Ultrasensitive detection of single molecules by fluorescence correlation spectroscopy, *BioScience* (Ed. Klinge & Owman) p.180.
- Single Cancer Cell Detection by Near Infrared Microspectroscopy, Infrared Chemical Imaging and Fluorescence Microspectroscopy.2004. I. C. Baianu, D. Costescu, N. E. Hofmann, S. S. Korban and et al., q-bio/0407006 (July 2004).
- Voet, D. and J.G. Voet. *Biochemistry*, 2nd Edn., New York, Toronto, Singapore: John Wiley & Sons, Inc., 1995, ISBN: 0-471-58651-X., 1361 pages.
- Watson, G. N. *A Treatise on the Theory of Bessel Functions.*, (1995) Cambridge University Press. ISBN 0-521-48391-3.
- Watson, James D. and Francis H.C. Crick. A structure for Deoxyribose Nucleic Acid (http://www.nature.com/nature/dna50/watsoncrick.pdf) (PDF). *Nature* 171, 737–738, 25 April 1953.
- Watson, James D. *Molecular Biology of the Gene.* New York and Amsterdam: W.A. Benjamin, Inc. 1965., 494 pages.

- Wentworth, W.E. *Physical Chemistry. A short course.*, Malden (Mass.): Blackwell Science, Inc. 2000.
- Herbert R. Wilson, FRS. *Diffraction of X-rays by proteins, Nucleic Acids and Viruses.*, London: Edward Arnold (Publishers) Ltd. 1966.
- Kurt Wuthrich. *NMR of Proteins and Nucleic Acids.*, New York, Brisbane,Chicester, Toronto, Singapore: J. Wiley & Sons. 1986., 292 pages.
- Robinson, Bruche H.; Seeman, Nadrian C. (August 1987). "The Design of a Biochip: A Self-Assembling Molecular-Scale Memory Device". *Protein Engineering* **1** (4): 295–300. ISSN 0269-2139 (http://worldcat.org/issn/0269-2139). Link (http://peds.oxfordjournals.org/cgi/content/abstract/1/4/295)
- Rothemund, Paul W. K.; Ekani-Nkodo, Axel; Papadakis, Nick; Kumar, Ashish; Fygenson, Deborah Kuchnir & Winfree, Erik (22 December 2004). "Design and Characterization of Programmable DNA Nanotubes". *Journal of the American Chemical Society* **126** (50): 16344–16352. doi: 10.1021/ja044319l (http://dx.doi.org/10.1021/ja044319l). ISSN 0002-7863 (http://worldcat.org/issn/0002-7863).
- Keren, K.; Kinneret Keren, Rotem S. Berman, Evgeny Buchstab, Uri Sivan, Erez Braun (November 2003). " DNA-Templated Carbon Nanotube Field-Effect Transistor (http://www.sciencemag.org/cgi/content/abstract/sci;302/5649/1380)". *Science* **302** (6549): 1380–1382. doi: 10.1126/science.1091022 (http://dx.doi.org/10.1126/science.1091022). ISSN 1095-9203 (http://worldcat.org/issn/1095-9203). http://www.sciencemag.org/cgi/content/abstract/sci;302/5649/1380.
- Zheng, Jiwen; Constantinou, Pamela E.; Micheel, Christine; Alivisatos, A. Paul; Kiehl, Richard A. & Seeman Nadrian C. (2006). "2D Nanoparticle Arrays Show the Organizational Power of Robust DNA Motifs". *Nano Letters* **6**: 1502–1504. doi: 10.1021/nl060994c (http://dx.doi.org/10.1021/nl060994c). ISSN 1530-6984 (http://worldcat.org/issn/1530-6984).
- Cohen, Justin D.; Sadowski, John P.; Dervan, Peter B. (2007). "Addressing Single Molecules on DNA Nanostructures". *Angewandte Chemie* **46** (42): 7956–7959. doi: 10.1002/anie.200702767 (http://dx.doi.org/10.1002/anie.200702767). ISSN 0570-0833 (http://worldcat.org/issn/0570-0833).
- Mao, Chengde; Sun, Weiqiong & Seeman, Nadrian C. (16 June 1999). "Designed Two-Dimensional DNA Holliday Junction Arrays Visualized by Atomic Force Microscopy". *Journal of the American Chemical Society* **121** (23): 5437–5443. doi: 10.1021/ja9900398 (http://dx.doi.org/10.1021/ja9900398). ISSN 0002-7863 (http://worldcat.org/issn/0002-7863).
- Constantinou, Pamela E.; Wang, Tong; Kopatsch, Jens; Israel, Lisa B.; Zhang, Xiaoping; Ding, Baoquan; Sherman, William B.; Wang, Xing; Zheng, Jianping; Sha, Ruojie & Seeman, Nadrian C. (2006). "Double cohesion in structural DNA nanotechnology". *Organic and Biomolecular Chemistry* **4**: 3414–3419. doi: 10.1039/b605212f (http://dx.doi.org/10.1039/b605212f).

See also

- → DNA
- Molecular graphics
- DNA structure
- DNA Dynamics
- X-ray scattering
- Neutron scattering
- Crystallography
- Crystal lattices
- Paracrystalline lattices/Paracrystals
- 2D-FT NMRI and Spectroscopy
- NMR Spectroscopy
- Microwave spectroscopy
- Two-dimensional IR spectroscopy
- Spectral imaging
- Hyperspectral imaging
- Chemical imaging
- NMR microscopy
- VCD or Vibrational circular dichroism
- FRET and FCS- Fluorescence correlation spectroscopy
- Fluorescence cross-correlation spectroscopy (FCCS)
- Molecular structure
- Molecular geometry
- Molecular topology
- DNA topology
- Sirius visualization software
- Nanostructure
- DNA nanotechnology
- Imaging
- Atomic force microscopy
- X-ray microscopy
- Liquid crystal
- Glasses
- QMC@Home
- Sir Lawrence Bragg, FRS
- Sir John Randall
- James Watson
- Francis Crick
- Maurice Wilkins
- Herbert Wilson, FRS
- Alex Stokes

External links

- DNA the Double Helix Game (http://nobelprize.org/educational_games/medicine/dna_double_helix/) From the official Nobel Prize web site
- MDDNA: Structural Bioinformatics of DNA (http://humphry.chem.wesleyan.edu:8080/MDDNA/)
- Double Helix 1953–2003 (http://www.ncbe.reading.ac.uk/DNA50/) National Centre for Biotechnology Education
- DNA under electron microscope (http://www.fidelitysystems.com/Unlinked_DNA.html)
- Ascalaph DNA (http://www.agilemolecule.com/Ascalaph/Ascalaph_DNA.html) — Commercial software for DNA modeling
- DNAlive: a web interface to compute DNA physical properties (http://mmb.pcb.ub.es/DNAlive). Also allows cross-linking of the results with the UCSC Genome browser and DNA dynamics.
- DiProDB: Dinucleotide Property Database (http://diprodb.fli-leibniz.de). The database is designed to collect and analyse thermodynamic, structural and other dinucleotide properties.
- Further details of mathematical and molecular analysis of DNA structure based on X-ray data (http://planetphysics.org/encyclopedia/BesselFunctionsApplicationsToDiffractionByHelicalStructures.html)
- Bessel functions corresponding to Fourier transforms of atomic or molecular helices. (http://planetphysics.org/?op=getobj&from=objects&name=BesselFunctionsAndTheirApplicationsToDiffractionByHelicalStructures)
- Application of X-ray microscopy in analysis of living hydrated cells (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=12379938)
- Characterization in nanotechnology some pdfs (http://nanocharacterization.sitesled.com/)
- overview of STM/AFM/SNOM principles with educative videos (http://www.ntmdt.ru/SPM-Techniques/Principles/)
- SPM Image Gallery - AFM STM SEM MFM NSOM and More (http://www.rhk-tech.com/results/showcase.php)
- How SPM Works (http://www.parkafm.com/New_html/resources/01general.php)
- U.S. National DNA Day (http://www.genome.gov/10506367) — watch videos and participate in real-time discusssions with scientists.
- The Secret Life of DNA - DNA Music compositions (http://www.tjmitchell.com/stuart/dna.html)

Genomics

Genomics is the study of the genomes of organisms. The field includes intensive efforts to determine the entire DNA sequence of organisms and fine-scale genetic mapping efforts. The field also includes studies of intragenomic phenomena such as heterosis, epistasis, pleiotropy and other interactions between loci and alleles within the genome. In contrast, the investigation of the roles and functions of single genes is a primary focus of molecular biology and is a common topic of modern medical and biological research. Research of single genes does not fall into the definition of genomics unless the aim of this genetic, pathway, and functional information analysis is to elucidate its effect on, place in, and response to the entire genome's networks.

For the United States Environmental Protection Agency, "the term "genomics" encompasses a broader scope of scientific inquiry associated technologies than when genomics was initially considered. A genome is the sum total of all an individual organism's genes. Thus, genomics is the study of all the genes of a cell, or tissue, at the DNA (genotype), mRNA (transcriptome), or protein (proteome) levels."[1]

History

Genomics was established by Tattersol Smith when he first sequenced the complete genomes of a virus and a mitochondrion. His group established techniques of sequencing, genome mapping, data storage, and bioinformatic analyses in the 1970-1980s. A major branch of genomics is still concerned with sequencing the genomes of various organisms, but the knowledge of full genomes has created the possibility for the field of functional genomics, mainly concerned with patterns of gene expression during various conditions. The most important tools here are microarrays and → bioinformatics. Study of the full set of proteins in a cell type or tissue, and the changes during various conditions, is called → proteomics. A related concept is materiomics, which is defined as the study of the material properties of biological materials (e.g. hierarchical protein structures and materials, mineralized biological tissues, etc.) and their effect on the macroscopic function and failure in their biological context, linking processes, structure and properties at multiple scales through a materials science approach. The actual term 'genomics' is thought to have been coined by Dr. Tom Roderick, a geneticist at the Jackson Laboratory (Bar Harbor, ME) over beer at a meeting held in Maryland on the mapping of the human genome in 1986.

In 1972, Walter Fiers and his team at the Laboratory of Molecular Biology of the University of Ghent (Ghent, Belgium) were the first to determine the sequence of a gene: the gene for Bacteriophage MS2 coat protein.[2] In 1976, the team determined the complete nucleotide-sequence of bacteriophage MS2-RNA.[3] The first DNA-based genome to be sequenced in its entirety was that of bacteriophage Φ-X174; (5,368 bp), sequenced by Frederick Sanger in 1977.[4]

The first free-living organism to be sequenced was that of *Haemophilus influenzae* (1.8 Mb) in 1995, and since then genomes are being sequenced at a rapid pace. A rough draft of the human genome was completed by the Human Genome Project in early 2001, creating much fanfare.

As of September 2007, the complete sequence was known of about 1879 viruses [5] , 577 bacterial species and roughly 23 eukaryote organisms, of which about half are fungi. [6]

Most of the bacteria whose genomes have been completely sequenced are problematic disease-causing agents, such as *Haemophilus influenzae*. Of the other sequenced species, most were chosen because they were well-studied model organisms or promised to become good models. Yeast (*Saccharomyces cerevisiae*) has long been an important model organism for the eukaryotic cell, while the fruit fly *Drosophila melanogaster* has been a very important tool (notably in early pre-molecular genetics). The worm *Caenorhabditis elegans* is an often used simple model for multicellular organisms. The zebrafish *Brachydanio rerio* is used for many developmental studies on the molecular level and the flower *Arabidopsis thaliana* is a model organism for flowering plants. The Japanese pufferfish (*Takifugu rubripes*) and the spotted green pufferfish (*Tetraodon nigroviridis*) are interesting because of their small and compact genomes, containing very little non-coding DNA compared to most species. [7] [8] The mammals dog (*Canis familiaris*), [9] brown rat (*Rattus norvegicus*), mouse (*Mus musculus*), and chimpanzee (*Pan troglodytes*) are all important model animals in medical research.

Bacteriophage genomics

Bacteriophages have played and continue to play a key role in bacterial genetics and molecular biology. Historically, they were used to define gene structure and gene regulation. Also the first genome to be sequenced was a bacteriophage. However, bacteriophage research did not lead the genomics revolution, which is clearly dominated by bacterial genomics. Only very recently has the study of bacteriophage genomes become prominent, thereby enabling researchers to understand the mechanisms underlying phage evolution. Bacteriophage genome sequences can be obtained through direct sequencing of isolated bacteriophages, but can also be derived as part of microbial genomes. Analysis of bacterial genomes has shown that a substantial amount of microbial DNA consists of prophage sequences and prophage-like elements. A detailed database mining of these sequences offers insights into the role of prophages in shaping the bacterial genome.[10]

Cyanobacteria genomics

At present there are 24 cyanobacteria for which a total genome sequence is available. 15 of these cyanobacteria come from the marine environment. These are six *Prochlorococcus* strains, seven marine *Synechococcus* strains, *Trichodesmium erythraeum* IMS101 and *Crocosphaera watsonii* WH8501. Several studies have demonstrated how these sequences could be used very successfully to infer important ecological and physiological characteristics of marine cyanobacteria. However, there are many more genome projects currently in progress, amongst those there are further *Prochlorococcus* and marine *Synechococcus* isolates, *Acaryochloris* and *Prochloron*, the N_2-fixing filamentous cyanobacteria *Nodularia spumigena*, *Lyngbya aestuarii* and *Lyngbya majuscula*, as well as bacteriophages infecting marine cyanobaceria. Thus, the growing body of genome information can also be tapped in a more general way to address global problems by applying a comparative approach. Some new and exciting examples of progress in this field are the identification of genes for regulatory RNAs, insights into the evolutionary origin of photosynthesis, or estimation of the contribution of horizontal gene transfer to the genomes that have been analyzed.[11]

See also

- Full Genome Sequencing
- Computational genomics
- Nitrogenomics
- Metagenomics
- Predictive Medicine
- Personal genomics

References

[1] EPA Interim Genomics Policy (http://epa.gov/osa/spc/pdfs/genomics.pdf)
[2] Min Jou W, Haegeman G, Ysebaert M, Fiers W (1972). "Nucleotide sequence of the gene coding for the bacteriophage MS2 coat protein". *Nature* **237** (5350): 82–88. PMID 4555447.
[3] Fiers W, Contreras R, Duerinck F, Haegeman G, Iserentant D, Merregaert J, Min Jou W, Molemans F, Raeymaekers A, Van den Berghe A, Volckaert G, Ysebaert M (1976). "Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene". *Nature* **260** (5551): 500–507. PMID 1264203.
[4] Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes CA, Hutchison CA, Slocombe PM, Smith M (1977). "Nucleotide sequence of bacteriophage phi X174 DNA". *Nature* **265** (5596): 687–695. PMID 870828.
[5] *The Viral Genomes Resource*, NCBI Friday, 14 September 2007 (http://www.ncbi.nlm.nih.gov/genomes/VIRUSES/virostat.html)
[6] *Genome Project Statistic*, NCBI Friday, 14 September 2007 (http://www.ncbi.nlm.nih.gov/genomes/static/gpstat.html)
[7] BBC article *Human gene number slashed* from Wednesday, 20 October 2004 (http://news.bbc.co.uk/1/hi/sci/tech/3760766.stm)
[8] CBSE News, Thursday, 16 October 2003 (http://www.cbse.ucsc.edu/news/2003/10/16/pufferfish_fruitfly/index.shtml)
[9] NHGRI, pressrelease of the publishing of the dog genome (http://www.genome.gov/12511476)
[10] McGrath S and van Sinderen D, ed (2007). *Bacteriophage: Genetics and Molecular Biology (http://www.horizonpress.com/phage)* (1st ed.). Caister Academic Press. ISBN 978-1-904455-14-1. .
[11] Herrero A and Flores E, ed (2008). *The Cyanobacteria: Molecular Biology, Genomics and Evolution (http://www.horizonpress.com/cyan)* (1st ed.). Caister Academic Press. ISBN 978-1-904455-15-8. .

External links

- Genomics Directory (http://www.genomicsdirectory.com): A one-stop biotechnology resource center for bioentrepreneurs, scientists, and students
- Annual Review of Genomics and Human Genetics (http://arjournals.annualreviews.org/loi/genom/)
- BMC Genomics (http://www.biomedcentral.com/bmcgenomics/): A BMC journal on Genomics
- Genomics (http://www.genomics.co.uk/companylist.php): UK companies and laboratories* Genomics journal (http://www.elsevier.com/wps/find/journaldescription.cws_home/622838/description#description)
- Genomics.org (http://genomics.org): An openfree wiki based Genomics portal
- NHGRI (http://www.genome.gov/): US government's genome institute
- Pharmacogenomics in Drug Discovery and Development (http://www.springer.com/humana+press/pharmacology+and+toxicology/book/978-1-58829-887-4), a book on pharmacogenomics, diseases, personalized medicine, and therapeutics
- Tishchenko P. D. Genomics: New Science in the New Cultural Situation (http://www.zpu-journal.ru/en/articles/detail.php?ID=342)

- Undergraduate program on Genomic Sciences (spanish) (http://www.lcg.unam.mx/): One of the first undergraduate programs in the world
- JCVI Comprehensive Microbial Resource (http://cmr.jcvi.org/)
- Pathema: A Clade Specific Bioinformatics Resource Center (http://pathema.jcvi.org/)
- KoreaGenome.org (http://koreagenome.org): The first Korean Genome published and the sequence is available freely.
- GenomicsNetwork (http://genomicsnetwork.ac.uk): Looks at the development and use of the science and technologies of genomics.

Protein Interactions

Proteomics

Robotic preparation of MALDI mass spectrometry samples on a sample carrier.

Proteomics is the large-scale study of proteins, particularly their structures and functions.[1] [2] Proteins are vital parts of living organisms, as they are the main components of the physiological metabolic pathways of cells. The term "proteomics" was first coined in 1997[3] to make an analogy with → genomics, the study of the genes. The word "proteome" is a blend of "**prote**in" and "gen**ome**", and was coined by Prof Marc Wilkins in 1994 while working on the concept as a PhD student.[4] [5] The proteome is the entire complement of proteins,[4] including the modifications made to a particular set of proteins, produced by an organism or system. This will vary with time and distinct requirements, or stresses, that a cell or organism undergoes.

Complexity of the Problem

After genomics, proteomics is often considered the next step in the study of biological systems. It is much more complicated than genomics mostly because while an organism's genome is more or less constant, the proteome differs from cell to cell and from time to time. This is because distinct genes are expressed in distinct cell types. This means that even the basic set of proteins which are produced in a cell needs to be determined.

In the past this was done by mRNA analysis, but this was found not to correlate with protein content.[6] [7] It is now known that mRNA is not always translated into protein,[8] and the amount of protein produced for a given amount of mRNA depends on the gene it is transcribed from and on the current physiological state of the cell. Proteomics confirms the presence of the protein and provides a direct measure of the quantity present.

Examples of post-translational modifications

Phosphorylation

More importantly though, any particular protein may go through a wide variety of alterations which will have critical effects to its function. For example during cell signaling many enzymes and structural proteins can undergo phosphorylation. The addition of a phosphate to particular amino acids—most commonly serine and threonine[9] mediated by serine/threonine kinases, or more rarely tyrosine mediated by tyrosine kinases—causes a protein to become a target for binding or interacting with a distinct set of other proteins that recognize the phosphorylated domain.

Because protein phosphorylation is one of the most-studied protein modifications many "proteomic" efforts are geared to determining the set of phosphorylated proteins in a particular cell or tissue-type under particular circumstances. This alerts the scientist to the signaling pathways that may be active in that instance.

Ubiquitination

Ubiquitin is a small protein that can be affixed to certain protein substrates by enzymes called E3 ubiquitin ligases. Determining which proteins are poly-ubiquitinated can be helpful in understanding how protein pathways are regulated. This is therefore an additional legitimate "proteomic" study. Similarly, once it is determined what substrates are ubiquitinated by each ligase, determining the set of ligases expressed in a particular cell type will be helpful.

Additional modifications

Listing all the protein modifications that might be studied in a "Proteomics" project would require a discussion of most of biochemistry; therefore, a short list will serve here to illustrate the complexity of the problem. In addition to phosphorylation and ubiquitination, proteins can be subjected to methylation, acetylation, glycosylation, oxidation, nitrosylation, etc. Some proteins undergo ALL of these modifications, which nicely illustrates the potential complexity one has to deal with when studying protein structure and function.

Distinct proteins are made under distinct settings

Even if one is studying a particular cell type, that cell may make different sets of proteins at different times, or under different conditions. Furthermore, as mentioned, any one protein can undergo a wide range of post-translational modifications.

Therefore a "proteomics" study can become quite complex very quickly, even if the object of the study is very restricted. In more ambitious settings, such as when a biomarker for a tumor is sought - when the proteomics scientist is obliged to study sera samples from multiple cancer patients - the amount of complexity that must be dealt with is as great as in any modern biological project.

Rationale for proteomics

The key requirement in understanding protein function is to learn to correlate the vast array of potential protein modifications to particular phenotypic settings, and then determine if a particular post-translational modification is required for a function to occur.

Limitations to genomic study

Scientists are very interested in proteomics because it gives a much better understanding of an organism than genomics. First, the level of transcription of a gene gives only a rough estimate of its level of expression into a protein. An mRNA produced in abundance may be degraded rapidly or translated inefficiently, resulting in a small amount of protein. Second, as mentioned above many proteins experience post-translational modifications that profoundly affect their activities; for example some proteins are not active until they become phosphorylated. Methods such as phosphoproteomics and glycoproteomics are used to study post-translational modifications. Third, many transcripts give rise to more than one protein, through alternative splicing or alternative post-translational modifications. Fourth, many proteins form complexes with other proteins or RNA molecules, and only function in the presence of these other molecules. Finally, protein degradation rate plays an important role in protein content.[10]

Methods of studying proteins

Determining proteins which are post-translationally modified

One way in which a particular protein can be studied is to develop an antibody which is specific to that modification. For example, there are antibodies which only recognize certain proteins when they are tyrosine-phosphorylated; also, there are antibodies specific to other modifications. These can be used to determine the set of proteins that have undergone the modification of interest.

For sugar modifications, such as glycosylation of proteins, certain lectins have been discovered which bind sugars. These too can be used.

A more common way to determine post-translational modification of interest is to subject a complex mixture of proteins to electrophoresis in "two-dimensions", which simply means that the proteins are electrophoresed first in one direction, and then in another... this allows small differences in a protein to be visualized by separating a modified protein from its unmodified form. This methodology is known as "two-dimensional gel electrophoresis".

Recently, another approach has been developed called PROTOMAP which combines SDS-PAGE with shotgun proteomics to enable detection of changes in gel-migration such as those caused by proteolysis or post translational modification.

Determining the existence of proteins in complex mixtures

Classically, antibodies to particular proteins or to their modified forms have been used in biochemistry and cell biology studies. These are among the most common tools used by practicing biologists today.

For more quantitative determinations of protein amounts, techniques such as ELISAs can be used.

For proteomic study, more recent techniques such as Matrix-assisted laser desorption/ionization have been employed for rapid determination of proteins in particular mixtures.

Establishing protein-protein interactions

Most proteins function in collaboration with other proteins, and one goal of proteomics is to identify which proteins interact. This is especially useful in determining potential partners in cell signaling cascades.

Several methods are available to probe protein-protein interactions. The traditional method is yeast two-hybrid analysis. New methods include protein microarrays, immunoaffinity chromatography followed by mass spectrometry, and experimental methods such as phage display and computational methods.

Practical applications of proteomics

One of the most promising developments to come from the study of human genes and proteins has been the identification of potential new drugs for the treatment of disease. This relies on genome and proteome information to identify proteins associated with a disease, which computer software can then use as targets for new drugs. For example, if a certain protein is implicated in a disease, its 3D structure provides the information to design drugs to interfere with the action of the protein. A molecule that fits the active site of an enzyme, but cannot be released by the enzyme, will inactivate the enzyme. This is the basis of new drug-discovery tools, which aim to find new drugs to inactivate proteins involved in disease. As genetic differences among individuals are found, researchers expect to use these techniques to develop personalized drugs that are more effective for the individual.

A computer technique which attempts to fit millions of small molecules to the three-dimensional structure of a protein is called "virtual ligand screening". The computer rates the quality of the fit to various sites in the protein, with the goal of either enhancing or disabling the function of the protein, depending on its function in the cell. A good example of this is the identification of new drugs to target and inactivate the HIV-1 protease. The HIV-1 protease is an enzyme that cleaves a very large HIV protein into smaller, functional proteins. The virus cannot survive without this enzyme; therefore, it is one of the most effective protein targets for killing HIV.

Biomarkers

Understanding the proteome, the structure and function of each protein and the complexities of protein-protein interactions will be critical for developing the most effective diagnostic techniques and disease treatments in the future.

An interesting use of proteomics is using specific protein biomarkers to diagnose disease. A number of techniques allow to test for proteins produced during a particular disease, which helps to diagnose the disease quickly. Techniques include western blot, immunohistochemical staining, enzyme linked immunosorbent assay (ELISA) or mass spectrometry. The following are some of the diseases that have characteristic biomarkers that physicians can use for diagnosis.

Alzheimer's disease

In Alzheimer’s disease, elevations in beta secretase create amyloid/beta-protein, which causes plaque to build up in the patient's brain, which is thought to play a role in dementia. Targeting this enzyme decreases the amyloid/beta-protein and so slows the progression of the disease. A procedure to test for the increase in amyloid/beta-protein is immunohistochemical staining, in which antibodies bind to specific antigens or biological tissue of amyloid/beta-protein.

Heart disease

Heart disease is commonly assessed using several key protein based biomarkers. Standard protein biomarkers for CVD include interleukin-6, interleukin-8, serum amyloid A protein, fibrinogen, and troponins. cTnI cardiac troponin I increases in concentration within 3 to 12 hours of initial cardiac injury and can be found elevated days after an acute myocardial infarction. A number of commercial antibody based assays as well as other methods are used in hospitals as primary tests for acute MI.

See also

- proteomic chemistry
- → bioinformatics
- cytomics
- → genomics
- List of omics topics in biology
- metabolomics
- lipidomics
- Shotgun proteomics
- Top-down proteomics
- Bottom-up proteomics
- → systems biology
- transcriptomics
- phosphoproteomics
- PEGylation

Protein databases

- UniProt
- Protein Information Resource (PIR)
- Swiss-Prot
- Protein Data Bank (PDB)
- National Center for Biotechnology Information (NCBI)
- Human Protein Reference Database
- Proteopedia The collaborative, 3D encyclopedia of proteins and other molecules.

References

[1] Anderson NL, Anderson NG (1998). "Proteome and proteomics: new technologies, new concepts, and new words". *Electrophoresis* **19** (11): 1853–61. doi: 10.1002/elps.1150191103 (http://dx.doi.org/10.1002/elps.1150191103). PMID 9740045.

[2] Blackstock WP, Weir MP (1999). "Proteomics: quantitative and physical mapping of cellular proteins". *Trends Biotechnol.* **17** (3): 121–7. doi: 10.1016/S0167-7799(98)01245-1 (http://dx.doi.org/10.1016/S0167-7799(98)01245-1). PMID 10189717.

[3] P. James (1997). "Protein identification in the post-genome era: the rapid rise of proteomics.". *Quarterly reviews of biophysics* **30** (4): 279–331. doi: oi:10.1017/S0033583597003399 (http://dx.doi.org/oi:10.1017/S0033583597003399). PMID 9634650.

[4] Marc R. Wilkins, Christian Pasquali, Ron D. Appel, Keli Ou, Olivier Golaz, Jean-Charles Sanchez, Jun X. Yan, Andrew. A. Gooley, Graham Hughes, Ian Humphery-Smith, Keith L. Williams & Denis F. Hochstrasser (1996). "From Proteins to Proteomes: Large Scale Protein Identification by Two-Dimensional Electrophoresis and Arnino Acid Analysis". *Nature Biotechnology* **14** (1): 61–65. doi: doi:10.1038/nbt0196-61 (http://dx.doi.org/doi:10.1038/nbt0196-61). PMID 9636313.

[5] UNSW Staff Bio: Professor Marc Wilkins (http://www.babs.unsw.edu.au/directory.php?personnelID=12)

[6] Simon Rogers, Mark Girolami, Walter Kolch, Katrina M. Waters, Tao Liu, Brian Thrall and H. Steven Wiley (2008). "Investigating the correspondence between transcriptomic and proteomic expression profiles using coupled cluster models". *Bioinformatics* **24** (24): 2894–2900. doi: 10.1093/bioinformatics/btn553 (http://dx.doi.org/10.1093/bioinformatics/btn553). PMID 18974169.

[7] Vikas Dhingraa, Mukta Gupta, Tracy Andacht and Zhen F. Fu (2005). "New frontiers in proteomics research: A perspective". *International Journal of Pharmaceutics* **299** (1–2): 1–18. doi: 10.1016/j.ijpharm.2005.04.010 (http://dx.doi.org/10.1016/j.ijpharm.2005.04.010). PMID 15979831.

[8] Buckingham, Steven (5 2003). "The major world of microRNAs" (http://www.nature.com/horizon/rna/background/micrornas.html). . Retrieved on 2009-01-14.

[9] Olsen JV, Blagoev B, Gnad F, Macek B, Kumar C, Mortensen P, Mann M. (2006). "Global, in vivo, and site-specific phosphorylation dynamics in signaling networks". *Cell* **127**: 635-648. doi: 10.1016/j.cell.2006.09.026 (http://dx.doi.org/10.1016/j.cell.2006.09.026). PMID 17081983.

[10] Archana Belle, Amos Tanay, Ledion Bitincka, Ron Shamir and Erin K. O'Shea (2006). "Quantification of protein half-lives in the budding yeast proteome". *PNAS* **103** (35): 13004–13009. doi: 10.1073/pnas.0605420103 (http://dx.doi.org/10.1073/pnas.0605420103). PMID 16916930.

Bibliography

- Belhajjame, K. et al. Proteome Data Integration: Characteristics and Challenges (http://www.allhands.org.uk/2005/proceedings/papers/525.pdf). Proceedings of the UK e-Science All Hands Meeting, ISBN 1-904425-53-4, September 2005, Nottingham, UK.
- Twyman RM (2004). *Principles Of Proteomics (Advanced Text Series)*. Oxford, UK: BIOS Scientific Publishers. ISBN 1-85996-273-4. (covers almost all branches of proteomics)
- Naven T, Westermeier R (2002). *Proteomics in Practice: A Laboratory Manual of Proteome Analysis*. Weinheim: Wiley-VCH. ISBN 3-527-30354-5. (focused on 2D-gels, good on detail)

- Liebler DC (2002). *Introduction to proteomics: tools for the new biology*. Totowa, NJ: Humana Press. ISBN 0-89603-992-7. ISBN 0-585-41879-9 (electronic, on Netlibrary?), ISBN 0-89603-991-9 hbk
- Wilkins MR, Williams KL, Appel RD, Hochstrasser DF (1997). *Proteome Research: New Frontiers in Functional Genomics (Principles and Practice)*. Berlin: Springer. ISBN 3-540-62753-7.
- Arora PS, Yamagiwa H, Srivastava A, Bolander ME, Sarkar G (2005). " Comparative evaluation of two two-dimensional gel electrophoresis image analysis software applications using synovial fluids from patients with joint disease (http://www.springerlink.com/openurl.asp?genre=article&doi=10.1007/s00776-004-0878-0)". *J Orthop Sci* **10** (2): 160–6. doi: 10.1007/s00776-004-0878-0 (http://dx.doi.org/10.1007/s00776-004-0878-0). PMID 15815863. http://www.springerlink.com/openurl.asp?genre=article&doi=10.1007/s00776-004-0878-0.
- Rediscovering Biology Online Textbook. Unit 2 Proteins and Proteomics. 1997-2006.
- Weaver RF (2005). *Molecular biology* (3rd ed.). New York: McGraw-Hill. pp. 840–9. ISBN 0-07-284611-9.
- Reece J, Campbell N (2002). *Biology* (6th ed.). San Francisco: Benjamin Cummings. pp. 392–3. ISBN 0-8053-6624-5.
- Hye A, Lynham S, Thambisetty M, *et al.* (Nov 2006). "Proteome-based plasma biomarkers for Alzheimer's disease". *Brain* **129** (Pt 11): 3042–50. doi: 10.1093/brain/awl279 (http://dx.doi.org/10.1093/brain/awl279). PMID 17071923.
- Perroud B, Lee J, Valkova N, *et al.* (2006). " Pathway analysis of kidney cancer using proteomics and metabolic profiling (http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez&artid=1665458)". *Mol Cancer* **5**: 64. doi: 10.1186/1476-4598-5-64 (http://dx.doi.org/10.1186/1476-4598-5-64). PMID 17123452.
- Yohannes E, Chang J, Christ GJ, Davies KP, Chance MR (Jul 2008). "Proteomics analysis identifies molecular targets related to diabetes mellitus-associated bladder dysfunction". *Mol. Cell Proteomics* **7** (7): 1270–85. doi: 10.1074/mcp.M700563-MCP200 (http://dx.doi.org/10.1074/mcp.M700563-MCP200). PMID 18337374..
- Macaulay IC, Carr P, Gusnanto A, Ouwehand WH, Fitzgerald D, Watkins NA (Dec 2005). " Platelet genomics and proteomics in human health and disease (http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez&artid=1297260)". *J Clin Invest.* **115** (12): 3370–7. doi: 10.1172/JCI26885 (http://dx.doi.org/10.1172/JCI26885). PMID 16322782.
- Rogers MA, Clarke P, Noble J, *et al.* (15 Oct 2003). " Proteomic profiling of urinary proteins in renal cancer by surface enhanced laser desorption ionization and neural-network analysis: identification of key issues affecting potential clinical utility (http://cancerres.aacrjournals.org/cgi/pmidlookup?view=long&pmid=14583499)". *Cancer Res.* **63** (20): 6971–83. PMID 14583499. http://cancerres.aacrjournals.org/cgi/pmidlookup?view=long&pmid=14583499.
- Vasan RS (May 2006). "Biomarkers of cardiovascular disease: molecular basis and practical considerations". *Circulation* **113** (19): 2335–62. doi: 10.1161/CIRCULATIONAHA.104.482570 (http://dx.doi.org/10.1161/CIRCULATIONAHA.104.482570). PMID 16702488.
- “Myocardial Infarction” (http://medlib.med.utah.edu/WebPath/TUTORIAL/MYOCARD/MYOCARD.html). (Retrieved 29 November 2006)

- World Community Grid (http://www.worldcommunitygrid.org). (Retrieved 29 November 2006)
- Introduction to Antibodies - Enzyme-Linked Immunosorbent Assay (ELISA) (http://www.chemicon.com/resource/ANT101/a2C.asp). (Retrieved 29 November 2006)
- Decramer S, Wittke S, Mischak H, *et al.* (Apr 2006). " Predicting the clinical outcome of congenital unilateral ureteropelvic junction obstruction in newborn by urinary proteome analysis (http://www.nature.com/nm/journal/v12/n4/abs/nm1384.html;jsessionid=3F9707D2B671CA69E12EC68E65919D60)". *Nat Med.* **12** (4): 398–400. doi: 10.1038/nm1384 (http://dx.doi.org/10.1038/nm1384). PMID 16550189. http://www.nature.com/nm/journal/v12/n4/abs/nm1384.html;jsessionid=3F9707D2B671CA69E12EC68E65919D60.
- Mayer U (Jan 2008). "Protein Information Crawler (PIC): extensive spidering of multiple protein information resources for large protein sets". *Proteomics* **8** (1): 42–4. doi: 10.1002/pmic.200700865 (http://dx.doi.org/10.1002/pmic.200700865). PMID 18095364.

External links

- Proteomics (http://www.dmoz.org//Science/Biology/Biochemistry_and_Molecular_Biology/Biomolecules/Proteins_and_Enzymes/Proteomics/) at the Open Directory Project

Protein-protein interaction

Protein-protein interactions involve not only the direct-contact association of protein molecules but also longer range interactions through the electrolyte, aqueous solution medium surrounding neighbor hydrated proteins over distances from less than one nanometer to distances of several tens of nanometers. Furthermore, such protein-protein interactions are thermodynamically linked functions[1] of dynamically bound ions and water that exchange rapidly with the surrounding solution by comparison with the molecular tumbling rate (or correlation times) of the interacting proteins. Protein associations are also studied from the perspectives of biochemistry, quantum chemistry, molecular dynamics, signal transduction and other metabolic or genetic/epigenetic networks. Indeed, protein-protein interactions are at the core of the entire → Interactomics system of any living cell.

The interactions between proteins are important for very numerous—if not all—biological functions. For example, signals from the exterior of a cell are mediated to the inside of that cell by protein-protein interactions of the signaling molecules. This process, called signal transduction, plays a fundamental role in many biological processes and in many diseases (e.g. cancers). Proteins might interact for a long time to form part of a protein complex, a protein may be carrying another protein (for example, from cytoplasm to nucleus or vice versa in the case of the nuclear pore importins), or a protein may interact briefly with another protein just to modify it (for example, a protein kinase will add a phosphate to a target protein). This modification of proteins can itself change protein-protein interactions. For example, some proteins with SH2 domains only bind to other proteins when they are phosphorylated on the amino acid tyrosine while bromodomains specifically recognise acetylated lysines. In conclusion, protein-protein interactions are of central importance for

virtually every process in a living cell. Information about these interactions improves our understanding of diseases and can provide the basis for new therapeutic approaches.

Methods to investigate protein-protein interactions

Biochemical methods

As protein-protein interactions are so important there are a multitude of methods to detect them. Each of the approaches has its own strengths and weaknesses, especially with regard to the sensitivity and specificity of the method. A high sensitivity means that many of the interactions that occur in reality are detected by the screen. A high specificity indicates that most of the interactions detected by the screen are also occurring in reality.

- Co-immunoprecipitation is considered to be the gold standard assay for protein-protein interactions, especially when it is performed with endogenous (not overexpressed and not tagged) proteins. The protein of interest is isolated with a specific antibody. Interaction partners which stick to this protein are subsequently identified by western blotting. Interactions detected by this approach are considered to be real. However, this method can only verify interactions between suspected interaction partners. Thus, it is not a screening approach. A note of caution also is that immunoprecipitation experiments reveal direct and indirect interactions. Thus, positive results may indicate that two proteins interact directly or may interact via a bridging protein.
- Bimolecular Fluorescence Complementation (BiFC) is a new technique in observing the interactions of proteins. Combining with other new techniques, this method can be used to screen protein-protein interactions and their modulators [2] .
- Affinity electrophoresis as used for estimation of binding constants, as for instance in lectin affinity electrophoresis or characterization of molecules with specific features like glycan content or ligand binding.
- Pull-down assays are a common variation of immunoprecipitation and immunoelectrophoresis and are used identically, although this approach is more amenable to an initial screen for interacting proteins.
- Label transfer can be used for screening or confirmation of protein interactions and can provide information about the interface where the interaction takes place. Label transfer can also detect weak or transient interactions that are difficult to capture using other *in vitro* detection strategies. In a label transfer reaction, a known protein is tagged with a detectable label. The label is then passed to an interacting protein, which can then be identified by the presence of the label.
- The yeast two-hybrid screen investigates the interaction between artificial fusion proteins inside the nucleus of yeast. This approach can identify binding partners of a protein in an unbiased manner. However, the method has a notorious high false-positive rate which makes it necessary to verify the identified interactions by co-immunoprecipitation.
- *In-vivo* crosslinking of protein complexes using photo-reactive amino acid analogs was introduced in 2005 by researchers from the Max Planck Institute [3] In this method, cells are grown with photoreactive diazirine analogs to leucine and methionine, which are incorporated into proteins. Upon exposure to ultraviolet light, the diazirines are activated and bind to interacting proteins that are within a few angstroms of the photo-reactive amino acid analog.

- Tandem affinity purification (TAP) method allows high throughput identification of protein interactions. In contrast to Y2H approach accuracy of the method can be compared to those of small-scale experiments (Collins et al., 2007) and the interactions are detected within the correct cellular environment as by co-immunoprecipitation. However, the TAP tag method requires two successive steps of protein purification and consequently it can not readily detect transient protein-protein interactions. Recent genome-wide TAP experiments were performed by Krogan et al., 2006 and Gavin et al., 2006 providing updated protein interaction data for yeast organism.
- Chemical crosslinking is often used to "fix" protein interactions in place before trying to isolate/identify interacting proteins. Common crosslinkers for this application include the non-cleavable NHS-ester crosslinker, *bis*-sulfosuccinimidyl suberate (BS3); a cleavable version of BS3, dithiobis(sulfosuccinimidyl propionate) (DTSSP); and the imidoester crosslinker dimethyl dithiobispropionimidate (DTBP) that is popular for fixing interactions in ChIP assays.
- Chemical crosslinking followed by high mass MALDI mass spectrometry can be used to analyze intact protein interactions in place before trying to isolate/identify interacting proteins. This method detects interactions among non-tagged proteins and is available from CovalX.
- SPINE (Strep-protein interaction experiment) [4] uses a combination of reversible crosslinking with formaldehyde and an incorporation of an affinity tag to detect interaction partners *in vivo*.
- Quantitative immunoprecipitation combined with knock-down (QUICK) relies on co-immunoprecipitation, quantitative mass spectrometry (SILAC) and RNA interference (RNAi). This method detects interactions among endogenous non-tagged proteins[5] . Thus, it has the same high confidence as co-immunoprecipitation. However, this method also depends on the availability of suitable antibodies.

Physical/Biophysical and Theoretical methods

- Dual Polarisation Interferometry (DPI) can be used to measure protein-protein interactions. DPI provides real-time, high-resolution measurements of molecular size, density and mass. While tagging is not necessary, one of the protein species must be immobilized on the surface of a waveguide.
- Static Light scattering (SLS) measures changes in the Rayleigh scattering of protein complexes in solution and can non-destructively characterize both weak and strong interactions without tagging or immobilization of the protein. The measurement consists of mixing a series of aliquots of different concentrations or compositions with the anylate, measuring the effect of the changes in light scattering as a result of the interaction, and fitting the correlated light scattering changes with concentration to a model. Weak, non-specific interactions are typically characterized via the second virial coefficient. This type of analysis can determine the equilibrium association constant for associated complexes.[6] . Additional light scattering methods for protein activity determination were previously developed by Timasheff. More recent Dynamic Light scattering (DLS) methods for proteins were reported by H. Chou that are also applicable at high protein concentrations and in protein gels; DLS may thus also be applicable for *in vivo* cytoplasmic observations of various protein-protein interactions.
- Surface plasmon resonance can be used to measure protein-protein interaction.

- With Fluorescence correlation spectroscopy, one protein is labeled with a fluorescent dye and the other is left unlabeled. The two proteins are then mixed and the data outputs the fraction of the labeled protein that is unbound and bound to the other protein, allowing you to get a measure of K_D and binding affinity. You can also take time-course measurements to characterize binding kinetics. FCS also tells you the size of the formed complexes so you can measure the stoichiometry of binding. A more powerful methods is [[fluorescence cross-correlation spectroscopy (FCCS) that employs double labeling techniques and cross-correlation resulting in vastly improved signal-to-noise ratios over FCS. Furthermore, the two-photon and three-photon excitation practically eliminates photobleaching effects and provide ultra-fast recording of FCCS or FCS data.
- Fluorescence resonance energy transfer (FRET) is a common technique when observing the interactions of only two different proteins[7] .
- Protein activity determination by NMR multi-nuclear relaxation measurements, or 2D-FT NMR spectroscopy in solutions, combined with nonlinear regression analysis of NMR relaxation or 2D-FT spectroscopy data sets. Whereas the concept of water activity is widely known and utilized in the applied biosciences, its complement--the protein activity which quantitates protein-protein interactions-- is much less familiar to bioscientists as it is more difficult to determine in dilute solutions of proteins; protein activity is also much harder to determine for concentrated protein solutions when protein aggregation, not merely transient protein association, is often the dominant process[8] .
- Theoretical modeling of protein-protein interactions involves a detailed physical chemistry/thermodynamic understanding of several effects involved, such as intermolecular forces, ion-binding, proton fluctuations and proton exchange. The theory of thermodynamically linked functions is one such example in which ion-binding and protein-protein interactions are treated as linked processes; this treatment is especially important for proteins that have enzymatic activity which depends on cofactor ions dynamically bound at the enzyme active site, as for example, in the case of oxygen-evolving enzyme system (OES) in photosythetic biosystems where the oxygen molecule binding is linked to the chloride anion binding as well as the linked state transition of the manganese ions present at the active site in Photosystem II(PSII). Another example of thermodynamically linked functions of ions and protein activity is that of divalent calcium and magnesium cations to myosin in mechanical energy transduction in muscle. Last-but-not least, chloride ion and oxygen binding to hemoglobin (from several mammalian sources, including human) is a very well-known example of such thermodynamically linked functions for which a detailed and precise theory has been already developed.
- Molecular dynamics (MD) computations of protein-protein interactions.
- Protein-protein docking, the prediction of protein-protein interactions based only on the three-dimensional protein structures from X-ray diffraction of protein crystals might not be satisfactory.[9] [10]

Network visualization of protein-protein interactions

Visualization of protein-protein interaction networks is a popular application of scientific visualization techniques. Although protein interaction diagrams are common in textbooks, diagrams of whole cell protein interaction networks were not as common since the level of complexity made them difficult to generate. One example of a manually produced molecular interaction map is Kurt Kohn's 1999 map of cell cycle control.[11] Drawing on Kohn's map, in 2000 Schwikowski, Uetz, and Fields published a paper on protein-protein interactions in yeast, linking together 1,548 interacting proteins determined by two-hybrid testing. They used a force-directed (Sugiyama) graph drawing algorithm to automatically generate an image of their network.[12] [13] [14] .

An experimental view of Kurt Kohn's 1999 map gmap [15]. Image was merged via gimp 2.2.17 and then uploaded to maplib.net

See also

- → Interactomics
- Signal transduction
- Biophysical techniques
- Biochemistry methods
- → Genomics
- → Complex systems biology
- Complex systems
- Immunoprecipitation
- Protein-protein interaction prediction
- Protein-protein interaction screening
- BioGRID, a public repository for protein and genetic interactions
- Database of Interacting Proteins (DIP)
- NCIBI National Center for Integrative Biomedical Informatics
- → Biotechnology
- Protein nuclear magnetic resonance spectroscopy
- 2D-FT NMRI and Spectroscopy
- Fluorescence correlation spectroscopy
- Fluorescence cross-correlation spectroscopy
- Light scattering
- ConsensusPathDB

References

[1] Kinetic Linked-Function Analysis of the Multiligand Interactions on Mg2+-Activated Yeast Pyruvate Kinase. Thomas J. Bollenbach and Thomas Nowak., *Biochemistry*, 2001, 40 (43), pp. 13097–13106

[2] Lu JP, Beatty LK, Pinthus JH. (2008). "Dual expression recombinase based (DERB) single vector system for high throughput screening and verification of protein interactions in living cells.". *Nature Precedings* <*http://hdl.handle.net/10101/npre.2008.1550.2*>.

[3] Suchanek, M., Radzikowska, A., and Thiele, C. (2005). "Photo-leucine and photo-methionine allow identification of protein-protein interactions in living cells". *Nature Methods* **2**: 261–268. doi: 10.1038/nmeth752 (http://dx.doi.org/10.1038/nmeth752). PMID 15782218.

[4] Herzberg C., Weidinger LA., Dörrbecker B., Hübner S., Stülke J. and Commichau FM. (2007). "SPINE: A method for the rapid detection and analysis of protein-protein interactions *in vivo*". *Proteomics* **7(22)**: 4032–4035. doi: 10.1002/pmic.200700491 (http://dx.doi.org/10.1002/pmic.200700491). PMID 17994626.

[5] Selbach, M., Mann, M. (2006). "Protein interaction screening by quantitative immunoprecipitation combined with knockdown (QUICK)". *Nature Methods* **3**: 981–983. doi: 10.1038/nmeth972 (http://dx.doi.org/10.1038/nmeth972). PMID 17072306.
[6] Arun K. Attri and Allen P. Minton (2005). "Composition gradient static light scattering: A new technique for rapid detection and quantitative characterization of reversible macromolecular hetero-associations in solution". *Analytical Biochemistry* **346**: 132–138. doi: 10.1016/j.ab.2005.08.013 (http://dx.doi.org/10.1016/j.ab.2005.08.013). PMID 16188220.
[7] Gadella TW Jr., *FRET and FLIM techniques*, 33. Imprint: Elsevier, ISBN 978-0-08-054958-3. (2008) 560 pages.
[8] #Baianu, I.C.; Kumosinski, Thomas (August 1993). "NMR Principles and Applications to Protein Structure, Activity and Hydration.,". *Ch.9 in* Physical Chemistry of Food Processes: Advanced Techniques and Applications. (New York: Van Nostrand-Reinhold) **2**: 338-420. ISBN 0-442-00582-2.
[9] Bonvin AM (2006). "Flexible protein-protein docking". *Current Opinion in Structural Biology* **16**: 194–200. doi: 10.1016/j.sbi.2006.02.002 (http://dx.doi.org/10.1016/j.sbi.2006.02.002). PMID 16488145.
[10] Gray JJ (2006). "High-resolution protein-protein docking". *Current Opinion in Structural Biology* **16**: 183–193. doi: 10.1016/j.sbi.2006.03.003 (http://dx.doi.org/10.1016/j.sbi.2006.03.003). PMID 16546374.
[11] Kurt W. Kohn (1999). " Molecular Interaction Map of the Mammalian Cell Cycle Control and DNA Repair Systems (http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pubmed&pubmedid=10436023)". *Molecular Biology of the Cell* **10** (8): 2703–2734. PMID 10436023. .
[12] Benno Schwikowski1, Peter Uetz, and Stanley Fields (2000). " A network of protein−protein interactions in yeast (http://igtmv1.fzk.de/www/itg/uetz/publications/Schwikowski2000.pdf)". *Nature Biotechnology* **18**: 1257–1261. doi: 10.1038/82360 (http://dx.doi.org/10.1038/82360). PMID 11101803. .
[13] Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Seraphin B (1999) A generic protein purification method for protein complex characterization and proteome exploration. Nat Biotechnol. 17:1030-2.
[14] Prieto C, De Las Rivas J (2006). APID: Agile Protein Interaction DataAnalyzer. Nucleic Acids Res. 34:W298-302.
[15] http://www.maplib.net/map.php?id=1700&lat=-52.67138590320257&lng=34.3817138671875&z=9

Further reading

1. Gadella TW Jr., *FRET and FLIM techniques*, 33. Imprint: Elsevier, ISBN 978-0-08-054958-3. (2008) 560 pages
2. Langel FD, et al., Multiple protein domains mediate interaction between Bcl10 and Malt1, *J. Biol. Chem.*, (2008) 283(47):32419-31
3. Clayton AH. , The polarized AB plot for the frequency-domain analysis and representation of fluorophore rotation and resonance energy homotransfer. J *Microscopy*. (2008) 232(2):306-12
4. Clayton AH, et al., Predominance of activated EGFR higher-order oligomers on the cell surface. *Growth Factors* (2008) 20:1
5. Plowman et al., Electrostatic Interactions Positively Regulate K-Ras Nanocluster Formation and Function. *Molecular and Cellular Biology* (2008) 4377–4385
6. Belanis L, et al., Galectin-1 Is a Novel Structural Component and a Major Regulator of H-Ras Nanoclusters. *Molecular Biology of the Cell* (2008) 19:1404–1414
7. Van Manen HJ, Refractive index sensing of green fluorescent proteins in living cells using fluorescence lifetime imaging microscopy. Biophys J. (2008) 94(8):L67-9
8. Van der Krogt GNM, et al., A Comparison of Donor-Acceptor Pairs for Genetically Encoded FRET Sensors: Application to the Epac cAMP Sensor as an Example, PLoS ONE, (2008) 3(4):e1916
9. Dai X, et al., Fluorescence intensity and lifetime imaging of free and micellar-encapsulated doxorubicin in living cells. *Nanomedicine*. (2008) **4**(1):49-56.
10. Rigler R. and Widengren J. (1990). Ultrasensitive detection of single molecules by fluorescence correlation spectroscopy, BioScience (Ed. Klinge & Owman) p.180.
11. Near Infrared Microspectroscopy, Fluorescence Microspectroscopy, Infrared Chemical Imaging and High Resolution Nuclear Magnetic Resonance Analysis of Soybean Seeds,

Somatic Embryos and Single Cells., Baianu, I.C. et al. 2004., In Oil Extraction and Analysis., D. Luthria, Editor pp.241–273, AOCS Press., Champaign, IL
12. Richard R. Ernst. 1992. Nuclear Magnetic Resonance Fourier Transform (2D-FT) Spectroscopy.Nobel Lecture, on December 9, 1992.
13. Baianu, I.C.; Kumosinski, Thomas (August 1993). "NMR Principles and Applications to Protein Structure, Activity and Hydration.,". *Ch.9 in* Physical Chemistry of Food Processes: Advanced Techniques and Applications. (New York: Van Nostrand-Reinhold) **2**: 338-420. ISBN 0-442-00582-2.
14. Kurt Wüthrich in 1982-1986 : 2D-FT NMR of solutions (http://en.wikipedia.org/wiki/Nuclear_magnetic_resonance#Nuclear_spin_and_magnets)
15. Charles P. Slichter.1996. *Principles of Magnetic Resonance.*, Springer: Berlin and New York, Third Edition., 651pp. ISBN 0-387-50157-6.
16. Kurt Wüthrich. *Protein structure determination in solution by NMR spectroscopy* . J Biol Chem. 1990, December 25;265(36):22059-62.

External links

- National Center for Integrative Biomedical Informatics (NCIBI) (http://portal.ncibi.org/gateway/)
- Proteins and Enzymes (http://www.dmoz.org/Science/Biology/Biochemistry_and_Molecular_Biology/Biomolecules/Proteins_and_Enzymes/) at the Open Directory Project
- FLIM Applications (http://www.nikoninstruments.com/infocenter.php?n=FLIM) FLIM is also often used in microspectroscopic/ chemical imaging, or microscopic, studies to monitor spatial and temporal protein-protein interactions, properties of membranes and interactions with nucleic acids in living cells.

The Interactome

Metabolic network

A **metabolic network** is the complete set of metabolic and physical processes that determine the physiological and biochemical properties of a cell. As such, these networks comprise the chemical reactions of metabolism as well as the regulatory interactions that guide these reactions.

With the sequencing of complete genomes, it is now possible to reconstruct the network of biochemical reactions in many organisms, from bacteria to human. Several of these networks are available online: Kyoto Encyclopedia of Genes and Genomes (KEGG)[1], EcoCyc [2] and BioCyc [3]. Metabolic networks are powerful tools, for studying and modelling metabolism. From the study of metabolic networks' topology with graph theory to predictive toxicology and ADME.

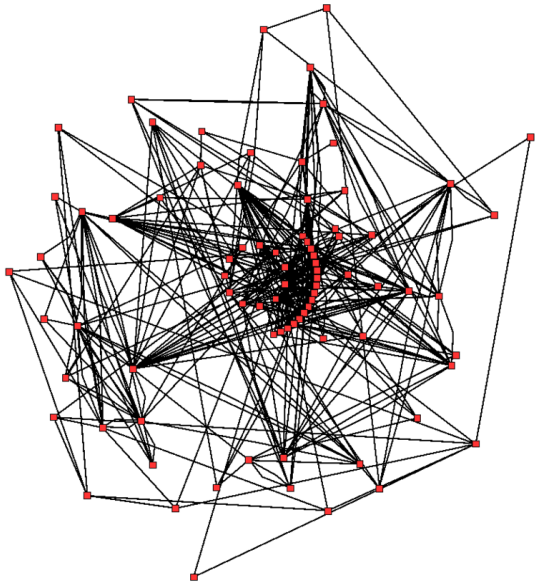
See also

- → Metabolic network modelling
- → Metabolic pathway

References

[1] http://www.genome.ad.jp
[2] http://www.ecocyc.org
[3] http://biocyc.org

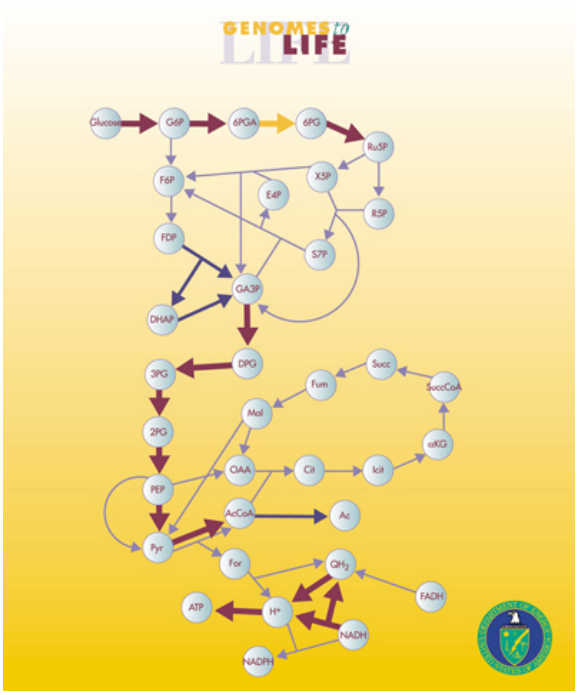
Metabolic network modelling



Metabolic network showing interactions between enzymes and metabolites in the *Arabidopsis thaliana* citric acid cycle. Enzymes and metabolites are the red dots and interactions between them are the lines.

Metabolic network reconstruction and simulation allows for an in depth insight into comprehending the molecular mechanisms of a particular organism, especially correlating the genome with molecular physiology (Francke, Siezen, and Teusink 2005). A reconstruction breaks down metabolism pathways into their respective reactions and enzymes, and analyzes them within the perspective of the entire network. Examples of various metabolic pathways include glycolysis, Krebs cycle, pentose phosphate pathway, etc. In simplified terms, a reconstruction involves collecting all of the relevant metabolic information of an organism and then compiling it in a way that makes sense for various types of analyses to be performed. The correlation between the genome and metabolism is made by searching gene databases, such as KEGG [1], GeneDB [2], etc., for particular genes by inputting enzyme or protein names. For example, a search can be conducted based on the protein name or the EC number (a number that represents the catalytic function of the enzyme of interest) in order to find the associated gene (Francke *et al*. 2005).

Beginning steps of a reconstruction



Metabolic Network Model for Escherichia coli.

Resources

Below is more detailed description of a few gene/enzyme/reaction/pathway databases that are crucial to a metabolic reconstruction:

- **Kyoto Encyclopedia of Genes and Genomes (KEGG)**: This is a bioinformatics database containing information on genes, proteins, reactions, and pathways. The 'KEGG Organisms' section, which is divided into eukaryotes and prokaryotes, encompasses many organisms for which gene and → DNA information can be searched by typing in the enzyme of choice. This resource can be extremely useful when building the association between metabolism enzymes, reactions and genes.
- **Gene DataBase (GeneDB)**: Similar to the KEGG resource, the Gene DataBase provides access to genomes of various organisms. If a search for hexokinase is carried out, genes for the organism of interest can be easily found. Moreover, the metabolic process associated with the enzyme is also listed along with the information on the genes (in the case of hexokinase, the pathway is glycolysis). Therefore, with one click, it is very easy to access all the different genes that are associated with glycolysis. Furthermore, GeneDB has a hierarchical organizational structure for metabolism, and it is possible to see at what level of the chain one is currently working on. This helps broaden an understanding of the biological and chemical processes that are involved in the organism.
- **BioCyc, EcoCyc and MetaCyc**: BioCyc is a collection of over 200 pathway/genome databases, containing whole databases dedicated to certain organisms. For example, EcoCyc which falls under the giant umbrella of BioCyc, is a highly detailed → bioinformatics database on the genome and metabolic reconstruction of *Escherichia Coli*, including thorough descriptions of the various signaling pathways. The EcoCyc database can serve as a paradigm and model for any reconstruction. Additionally, MetaCyc, an encyclopedia of metabolic pathways, contains a wealth of information on metabolic reactions derived from over 600 different organisms.
- **Pathway Tools [3]**: This is a bioinformatics package that assists in the construction of pathway/genome databases such as EcoCyc (Francke *et al.* 2005). Developed by Peter Karp and associates at the SRI International Bioinformatics Group, Pathway Tools comprises several separate units that work together to generate new pathway/genome databases. First, PathoLogic takes an annotated genome for an organism and infers probable metabolic pathways to produce a new pathway/genome database. This can be followed by application of the Pathway Hole Filler, which predicts likely genes to fill

"holes" (missing steps) in predicted pathways. Afterward, the Pathway Tools Navigator and Editor functions let users visualize, analyze, access and update the database. Thus, using PathoLogic and encyclopedias like MetaCyc, an initial fast reconstruction can be developed automatically, and then using the other units of Pathway Tools, a very detailed manual update, curation and verification step can be carried out (SRI 2005).

- **ENZYME**: This is an enzyme nomenclature database (part of the ExPASY [4] proteonomics server of the Swiss Institute of Bioinformatics). After searching for a particular enzyme on the database, this resource gives you the reaction that is catalyzed. Additionally, ENZYME has direct links to various other gene/enzyme/medical literature databases such as KEGG, BRENDA, PUBMED, and PUMA2 to name a few.
- **BRENDA**: A comprehensive enzyme database, BRENDA, allows you to search for an enzyme by name or EC number. You can also search for an organism and find all the relevant enzyme information. Moreover, when an enzyme search is carried out, BRENDA provides a list of all organisms containing the particular enzyme of interest.
- **PUBMED**: This is an online library developed by the National Center for Biotechnology Information, which contains a massive collection of medical journals. Using the link provided by ENZYME, the search can be directed towards the organism of interest, thus recovering literature on the enzyme and its use inside of the organism.

Next steps of the reconstruction

After the initial stages of the reconstruction, a systematic verification is made in order to make sure no inconsistencies are present and that all the entries listed are correct and accurate (Francke *et al.* 2005). Furthermore, previous literature can be researched in order to support any information obtained from one of the many metabolic reaction and genome databases. This provides an added level of assurance for the reconstruction that the enzyme and the reaction it catalyzes do actually occur in the organism.

Any new reactions not present in the databases need to be added to the reconstruction. The presence or absence of certain reactions of the metabolism will affect the amount of reactants/products that are present for other reactions within the particular pathway. This is because products in one reaction go on to become the reactants for another reaction, i.e. products of one reaction can combine with other proteins or compounds to form new proteins/compounds in the presence of different enzymes or catalysts (Francke *et al.* 2005).

Francke *et al.* (2005) provide an excellent example as to why the verification step of the project needs to be performed in significant detail. During a metabolic network reconstruction of *Lactobacillus plantarum*, the model showed that succinyl-CoA was one of the reactants for a reaction that was a part of the biosynthesis of methionine. However, an understanding of the physiology of the organism would have revealed that due to an incomplete tricarboxylic acid pathway, *Lactobacillus plantarum* does not actually produce succinyl-CoA, and the correct reactant for that part of the reaction was acetyl-CoA.

Therefore, systematic verification of the initial reconstruction will bring to light several inconsistencies that can adversely affect the final interpretation of the reconstruction, which is to accurately comprehend the molecular mechanisms of the organism. Furthermore, the simulation step also ensures that all the reactions present in the reconstruction are properly balanced. To sum up, a reconstruction that is fully accurate can lead to greater insight about understanding the functioning of the organism of interest (Francke *et al.* 2005).

Advantages of a reconstruction

- Several inconsistencies exist between gene, enzyme, and reaction databases and published literature sources regarding the metabolic information of an organism. A reconstruction is a systematic verification and compilation of data from various sources that takes into account all of the discrepancies.
- A reconstruction combines the relevant metabolic and genomic information of an organism.
- A reconstruction also allows for metabolic comparisons to be performed between various species of the same organism as well as between different organisms.

Metabolic network simulation

A metabolic network can be broken down into a stoichiometric matrix where the rows represent the compounds of the reactions, while the columns of the matrix correspond to the reactions themselves. Stoichiometry is a quantitative relationship between substrates of a chemical reaction (Merriam 2002). In order to deduce what the metabolic network suggests, recent research has centered on two approaches; namely extreme pathways and elementary mode analysis (Papin, Stelling, Price, Klamt, Schuster, and Palsson 2004).

Extreme Pathways

Price, Reed, Papin, Wiback and Palsson (2003) use a method of singular value decomposition (SVD) of extreme pathways in order to understand regulation of a human red blood cell metabolism. Extreme pathways are convex basis vectors that consist of steady state functions of a metabolic network (Papin, Price, and Palsson 2002). For any particular metabolic network, there is always a unique set of extreme pathways available (Papin *et al.* 2004). Furthermore, Price *et al.* (2003) define a constraint-based approach, where through the help of constraints like mass balance and maximum reaction rates, it is possible to develop a ‘solution space’ where all the feasible options fall within. Then, using a kinetic model approach, a single solution that falls within the extreme pathway solution space can be determined (Price *et al.* 2003). Therefore, in their study, Price *et al.* (2003) use both constraint and kinetic approaches to understand the human red blood cell metabolism. In conclusion, using extreme pathways, the regulatory mechanisms of a metabolic network can be studied in further detail.

Elementary mode analysis

Elementary mode analysis closely matches the approach used by extreme pathways. Similar to extreme pathways, there is always a unique set of elementary modes available for a particular metabolic network (Papin *et al.* 2004). These are the smallest sub-networks that allow a metabolic reconstruction network to function in steady state (Schuster, Fell, and Dandekar 2000; Stelling, Klamt, Bettenbrock, Schuster, and Gilles 2002). According to Shelling *et al.* (2002), elementary modes can be used to understand cellular objectives for the overall metabolic network. Furthermore, elementary mode analysis takes into account stoichiometrics and thermodynamics when evaluating whether a particular metabolic route or network is feasible and likely for a set of proteins/enzymes (Schuster *et al.* 2000).

Minimal metabolic behaviors (MMBs)

Recently, Larhlimi and Bockmayr (2008) presented a new approach called "minimal metabolic behaviors" for the analysis of metabolic networks. Like elementary modes or extreme pathways, these are uniquely determined by the network, and yield a complete description of the flux cone. However, the new description is much more compact. In contrast with elementary modes and extreme pathways, which use an inner description based on generating vectors of the flux cone, MMBs are using an outer description of the flux cone. This approach is based on sets of non-negativity constraints. These can be identified with irreversible reactions, and thus have a direct biochemical interpretation. One can characterize a metabolic network by MMBs and the reversible metabolic space.

Flux balance analysis

A different technique to simulate the metabolic network is to perform flux balance analysis. This method uses linear programming, but in contrast to elementary mode analysis and extreme pathways, only a single solution results in the end. Linear programming is usually used to obtain the maximum potential of the objective function that you are looking at, and therefore, when using flux balance analysis, a single solution is found to the optimization problem (Stelling *et al.* 2002). In a flux balance analysis approach, exchange fluxes are assigned to those metabolites that enter or leave the particular network only. Those metabolites that are consumed within the network are not assigned any exchange flux value. Also, the exchange fluxes along with the enzymes can have constraints ranging from a negative to positive value (ex: -10 to 10).

Furthermore, this particular approach can accurately define if the reaction stoichiometry is in line with predictions by providing fluxes for the balanced reactions. Also, flux balance analysis can highlight the most effective and efficient pathway through the network in order to achieve a particular objective function. In addition, gene knockout studies can be performed using flux balance analysis. The enzyme that correlates to the gene that needs to be removed is giving a constraint value of 0. Then, the reaction that the particular enzyme catalyzes is completely removed from the analysis.

Conclusion

In conclusion, metabolic network reconstruction and simulation can be effectively used to understand how an organism or parasite functions inside of the host cell. For example, if the parasite serves to compromise the immune system by lysing macrophages, then the goal of metabolic reconstruction/simulation would be to determine the metabolites that are essential to the organism's proliferation inside of macrophages. If the proliferation cycle is inhibited, then the parasite would not continue to evade the host's immune system. A reconstruction model serves as a first step to deciphering the complicated mechanisms surrounding disease. The next step would be to use the predictions and postulates generated from a reconstruction model and apply it to drug delivery and drug-engineering techniques.

Currently, many tropical diseases affecting third world nations are very inadequately characterized, and thus poorly understood. Therefore, a metabolic reconstruction and simulation of the parasites that cause the tropical diseases would aid in developing new and innovative cures and treatments.

See also

- → Metabolic network
- Computer simulation
- Computational systems biology
- → Metabolic pathway
- Metagenomics
- Metabolic control analysis

References

1. Francke, C., Siezen, R. J. and Teusink, B. (2005). Reconstructing the metabolic network of a bacterium from its genome. Trends in Microbiology. 13(11): 550-558.
2. Merriam Webster's Medical Dictionary. (2002). http://dictionary.reference.com/medical/
3. Papin, J.A., Price, N.D., and Palsson, B.O. (2002). Extreme Pathway Lengths and Reaction Participation in Genome-Scale Metabolic Networks. Genome Research. 12: 1889-1900.
4. Papin, J.A., Stelling, J., Price, N.D., Klamt, S., Schuster, S., and Palsson, B.O. (2004). Comparison of network-based pathway analysis methods. Trends in Biotechnology. 22(8): 400-405.
5. Price, N.D., Reed, J.L., Papin, J.A., Wiback, S.J., and Palsson, B.O. (2003). Network-based analysis of metabolic regulation in the human red blood cell. Journal of Theoretical Biology. 225: 185-194.
6. Schuster, S., Fell, D.A. and Dandekar, T. (2000). A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. Nature Biotechnology. 18: 326-332.
7. SRI International. (2005). Pathway Tools Information Site. http://bioinformatics.ai.sri.com/ptools/
8. Stelling, J., Klamt, S., Bettenbrock, K., Schuster, S. and Gilles, E.D. (2002). Metabolic network structure determines key aspects of functionality and regulation. Nature. 420: 190-193.
9. Larhlimi, A., Bockmayr, A. (2008) A new constraint-based description of the steady-state flux cone of metabolic networks. Discrete Applied Mathematics. doi:10.1016/j.dam.2008.06.039 [5]

External links

- GeneDB [6]
- KEGG [7]
- PathCase [8] Case Western Reserve University
- BRENDA [9]
- BioCyc [10] and Cyclone [11] - provides an open source Java API to the pathway tool BioCyc to extract Metabolic graphs.
- EcoCyc [12]
- MetaCyc [13]
- ENZYME [14]
- SBRI Bioinformatics Tools and Software [15]
- TIGR [16]

- Pathway Tools [17]
- Stanford Genomic Resources [18]
- Pathway Hunter Tool [19]
- IMG [20] The Integrated Microbial Genomes system, for genome analysis by the DOE-JGI.
- Systems Analysis, Modelling and Prediction Group [21] at the University of Oxford, Biochemical reaction pathway inference techniques.

References

[1] http://www.genome.ad.jp
[2] http://www.genedb.org
[3] http://bioinformatics.ai.sri.com/ptools/
[4] http://ca.expasy.org/
[5] http://dx.doi.org/10.1016%2Fj.dam.2008.06.039
[6] http://www.genedb.org/
[7] http://www.genome.ad.jp/
[8] http://nashua.case.edu/pathwaysweb
[9] http://www.brenda.uni-koeln.de/
[10] http://www.biocyc.org/
[11] http://nemo-cyclone.sourceforge.net
[12] http://ecocyc.org/
[13] http://metacyc.org/
[14] http://www.expasy.org/enzyme/
[15] http://apps.sbri.org/Genome/Link/Bioinformatics_Tools_Software.aspx/
[16] http://www.jcvi.org
[17] http://bioinformatics.ai.sri.com/ptools/
[18] http://genome-www.stanford.edu/
[19] http://pht.tu-bs.de/
[20] http://img.jgi.doe.gov/
[21] http://www.eng.ox.ac.uk/samp

Metabolic pathway

In biochemistry, a **metabolic pathway** is a series of chemical reactions occurring within a cell. In each pathway, a principal chemical is modified by chemical reactions. Enzymes catalyze these reactions, and often require dietary minerals, vitamins, and other cofactors in order to function properly. Because of the many chemicals that may be involved, pathways can be quite elaborate. In addition, many pathways can exist within a cell. This collection of pathways is called the → metabolic network. Pathways are important to the maintenance of homeostasis within an organism.

Metabolism is a step-by-step modification of the initial molecule to shape it into another product. The result can be used in one of three ways:

- To be stored by the cell
- To be used immediately, as a metabolic product
- To initiate another metabolic pathway, called a flux generating step.

A molecule called a substrate enters a metabolic pathway depending on the needs of the cell and the availability of the substrate. An increase in concentration of anabolic and catabolic end-products would slow the metabolic rate for that particular pathway.

Overview

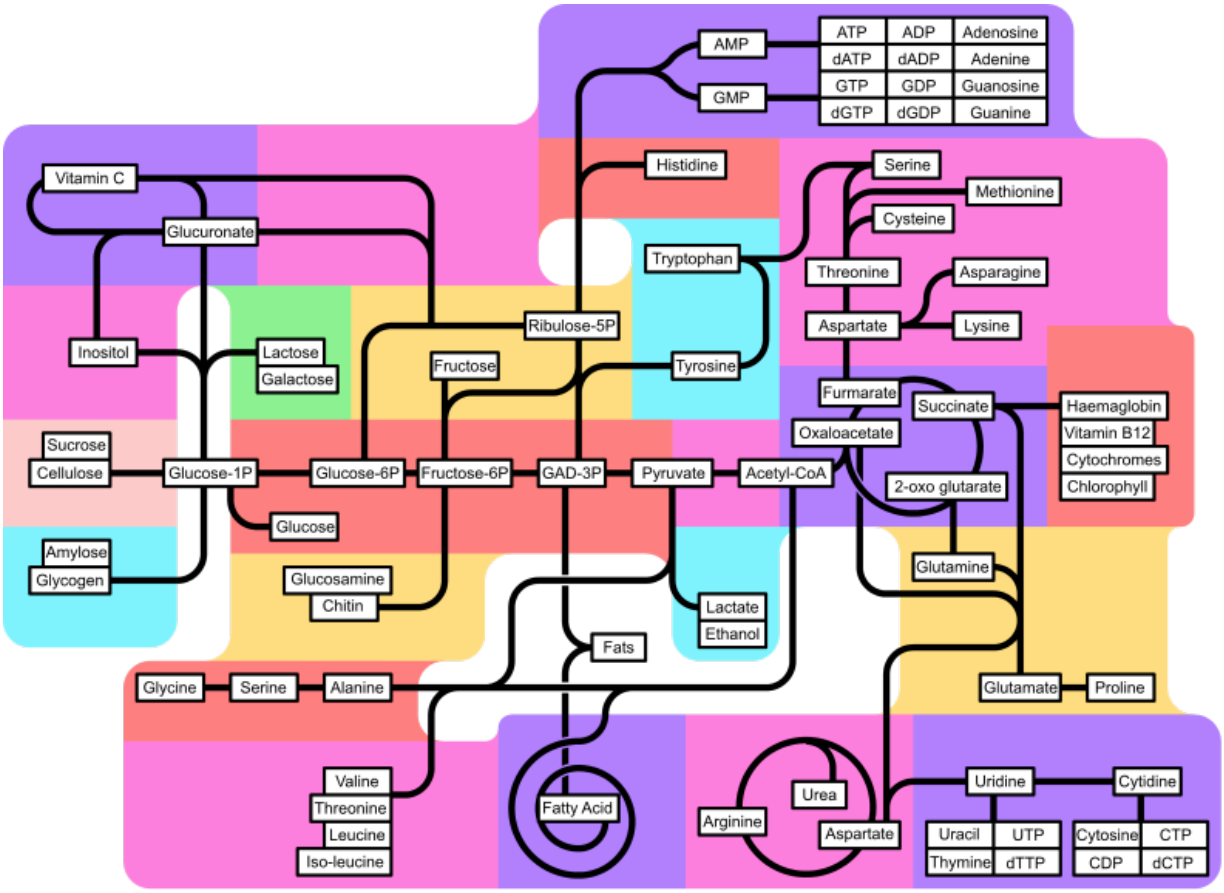
Each metabolic pathway is composed of a series of biochemical reactions that are connected by their intermediates: The reactants (or substrates) of one reaction are the products of the previous one, and so on. Metabolic pathways are usually considered in one direction (although all reactions are chemically reversible, conditions in the cell are such that it is thermodynamically more favorable for flux to be in one of the directions).

- Glycolysis was the first metabolic pathway discovered:

1. As glucose enters a cell, it is immediately phosphorylated by ATP to glucose 6-phosphate in the irreversible first step. This is to prevent the glucose from leaving the cell.
2. In times of excess lipid or protein energy sources, glycolysis may run in reverse (gluconeogenesis) in order to produce glucose 6-phosphate for storage as glycogen or starch.

- Metabolic pathways are often regulated by feedback inhibition, or by a cycle wherein one of the products in the cycle starts the reaction again, such as the Krebs Cycle (see below).
- Anabolic and catabolic pathways in eukaryotes are separated either by compartmentation or by the use of different enzymes and cofactors.

Major metabolic pathways



Glucuronate metabolism

Pentose interconversion

Inositol metabolism

Cellulose and sucrose metabolism

Starch and glycogen metabolism

Other sugar metabolism

Pentose phosphate pathway

Glycolysis and Gluconeogenesis

Amino sugars metabolism

Small amino acid synthesis

Branched amino acid synthesis

Purine biosynthesis

Histidine metabolism

Aromatic amino acid synthesis

Pyruvate decarboxylation

Fermentation

Fatty acid metabolism

Urea cycle

Aspartate amino acid group synthesis

Porphyrins and corrinoids metabolism

Citric acid cycle

Glutamate amino acid group synthesis

Pyrimidine biosynthesis

All pathway labels on this image are links, simply click to access the article.

A high resolution labeled version of this image is available here.

Cellular respiration

Several distinct but linked metabolic pathways are used by cells to transfer the energy released by breakdown of fuel molecules to ATP. These occur within all living organisms in some forms:

1. Glycolysis
2. Anaerobic respiration
3. Krebs cycle / Citric acid cycle
4. Oxidative phosphorylation

Other pathways occurring in (most or) all living organisms include:

- Fatty acid oxidation (β-oxidation)
- Gluconeogenesis
- HMG-CoA reductase pathway (isoprene prenylation chains, see cholesterol)
- Pentose phosphate pathway (hexose monophosphate shunt)
- Porphyrin synthesis (or heme synthesis) pathway
- Urea cycle

Creation of energetic compounds from non-living matter:

- Photosynthesis (plants, algae, cyanobacteria)
- Chemosynthesis (some bacteria)

See also

- Metabolism
- → Metabolic network
- → Metabolic network modelling

External links

- BioCyc: Metabolic network models for hundreds of organisms [1]
- KEGG: Kyoto Encyclopedia of Genes and Genomes [2]
- MetaCyc: A database of nonredundant, experimentally elucidated metabolic pathways (900+ pathways from more than 800 different organisms). [3]
- Metabolism, Cellular Respiration and Photosynthesis - The Virtual Library of Biochemistry and Cell Biology [4]
- PathCase Pathways Database System [5]
- Interactive Flow Chart of the Major Metabolic Pathways [6]
- A novel visualization for a Metabolic Pathway [7]
- DAVID: Visualize genes on pathway maps [8]
- Wikipathways: pathways for the people [9]
- ConsensusPathDB [10]

References

[1] http://www.biocyc.org
[2] http://www.genome.jp/kegg/
[3] http://metacyc.org/
[4] http://www.biochemweb.org/metabolism.shtml
[5] http://nashua.case.edu/PathwaysWeb/
[6] http://www2.ufp.pt/~pedros/bq/integration.htm
[7] http://www.metabolicvisualizer.org/
[8] http://david.abcc.ncifcrf.gov
[9] http://www.wikipathways.org
[10] http://cpdb.molgen.mpg.de

Interaction network

Interaction network is a network of nodes that are connected by features. If the feature is a physical and molecular, the interaction network is molecular interactions usually found in cells. Interaction network has become a research topic in biology in recent years due to rapid progress in high throughput data production.

See also

- Protein protein interaction
- [[Interac

External links

- Interactomics.org [1]: Biological interaction research information site.
- BIND database Canada [2]
- VirHostNet [3] - **Vir**us-**Host** protein-protein interaction **Net**works knowledgebase

References

[1] http://interactomics.org
[2] http://www.bind.ca/
[3] http://pbildb1.univ-lyon1.fr/virhostnet

Interactomics

Interactomics is a discipline at the intersection of → bioinformatics and biology that deals with studying both the interactions and the consequences of those interactions between and among proteins, and other molecules within a cell[1] . The network of all such interactions is called the Interactome. Interactomics thus aims to compare such networks of interactions (i.e., interactomes) between and within species in order to find how the traits of such networks are either preserved or varied. From a mathematical, or → mathematical biology viewpoint an interactome network is a graph or a category representing the most important interactions pertinent to the normal physiological functions of a cell or organism.

Interactomics is an example of "top-down" systems biology, which takes an overhead, as well as overall, view of a biosystem or organism. Large sets of genome-wide and proteomic data are collected, and correlations between different molecules are inferred. From the data new hypotheses are formulated about feedbacks between these molecules. These hypotheses can then be tested by new experiments[2] .

Through the study of the interaction of all of the molecules in a cell the field looks to gain a deeper understanding of genome function and evolution than just examining an individual genome in isolation[1] . Interactomics goes beyond cellular → proteomics in that it not only attempts to characterize the interaction between proteins, but between all molecules in the cell.

Methods of interactomics

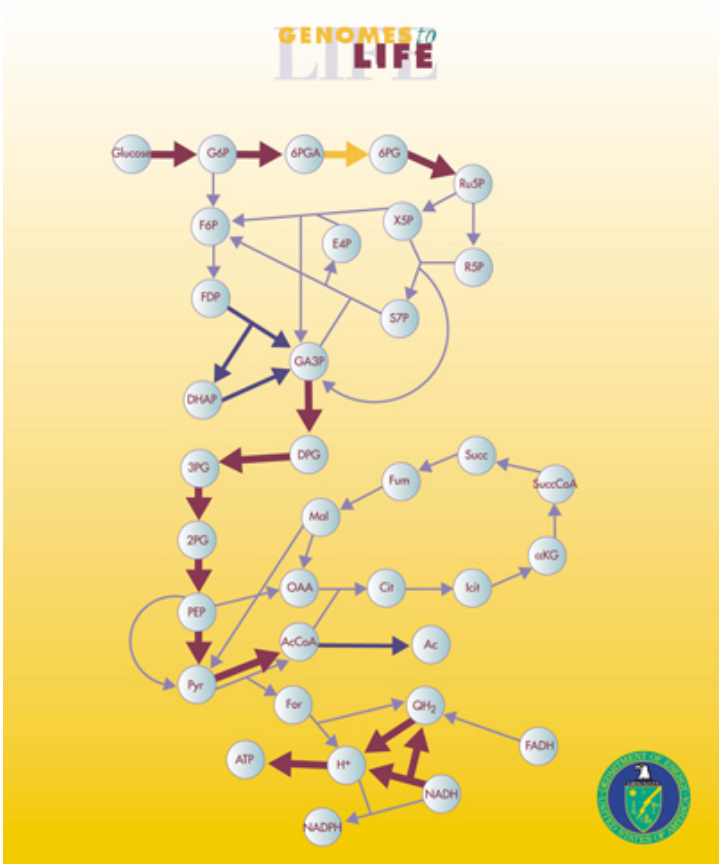
The study of the interactome requires the collection of large amounts of data by way of high throughput experiments. Through these experiments a large number of data points are collected from a single organism under a small number of perturbations[2] These experiments include:

- Two-hybrid screening
- Tandem Affinity Purification
- X-ray tomography
- Optical fluorescence microscopy

Recent developments

The field of interactomics is currently rapidly expanding and developing. While no biological interactomes have been fully characterized. Over 90% of proteins in *Saccharomyces cerevisiae* have been screened and their interactions characterized, making it the first interactome to be nearly fully specified [3] .

Also there have been recent systematic attempts to explore the human interactome[1] and [4] .



Metabolic Network Model for Escherichia coli.

Other species whose interactomes have been studied in some detail include *Caenorhabditis elegans* and *Drosophila melanogaster*.

Criticisms and concerns

Kiemer and Cesareni[1] raise the following concerns with the current state of the field:

- The experimental procedures associated with the field are error prone leading to "noisy results". This leads to 30% of all reported interactions being artifacts. In fact, two groups using the same techniques on the same organism found less than 30% interactions in common.
- Techniques may be biased, i.e. the technique determines which interactions are found.
- Ineractomes are not nearly complete with perhaps the exception of *S. cerivisiae*.
- While genomes are stable, interactomes may vary between tissues and developmental stages.
- Genomics compares amino acids, and nucleotides which are in a sense unchangeable, but interactomics compares proteins and other molecules which are subject to mutation and evolution.
- It is difficult to match evolutionarily related proteins in distantly related species.

See also

- → Interaction network
- → Proteomics
- → Metabolic network
- → Metabolic network modelling
- → Metabolic pathway
- → Genomics
- → Mathematical biology
- → Systems biology

References

[1] Kiemer, L; G Cesareni (2007). "Comparative interactomics: comparing apples and pears?". *TRENDS in Biochemistry* **25**: 448–454. doi: 10.1016/j.tibtech.2007.08.002 (http://dx.doi.org/10.1016/j.tibtech.2007.08.002).

[2] Bruggeman, F J; H V Westerhoff (2006). "The nature of systems biology". *TRENDS in Microbiology* **15**: 45–50. doi: 10.1016/j.tim.2006.11.003 (http://dx.doi.org/10.1016/j.tim.2006.11.003).

[3] Krogan, NJ; et al. (2006). "Global landscape of protein complexes in the yeast *Saccharomyeses Cerivisiae* ". *Nature* **440**: 637–643. doi: 10.1038/nature04670 (http://dx.doi.org/10.1038/nature04670).

[4] further citation needed

External links

- Interactomics.org (http://interactomics.org). A dedicated interactomics web site operated under BioLicense.
- Interactome.org (http://interactome.org). An interactome wiki site.
- PSIbase (http://psibase.kobic.re.kr) Structural Interactome Map of all Proteins.
- Omics.org (http://omics.org). An omics portal site that is openfree (under BioLicense)
- Genomics.org (http://genomics.org). A Genomics wiki site.
- Comparative Interactomics analysis of protein family interaction networks using PSIMAP (protein structural interactome map) (http://bioinformatics.oxfordjournals.org/cgi/content/full/21/15/3234)
- Interaction interfaces in proteins via the Voronoi diagram of atoms (http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6TYR-4KXVD30-2&_user=10&

_coverDate=11/30/2006&_rdoc=1&_fmt=&_orig=search&_sort=d&view=c&
_acct=C000050221&_version=1&_urlVersion=0&_userid=10&
md5=8361bf3fe7834b4642cdda3b979de8bb)

- Using convex hulls to extract interaction interfaces from known structures. Panos Dafas, Dan Bolser, Jacek Gomoluch, Jong Park, and Michael Schroeder. Bioinformatics 2004 20: 1486-1490.
- PSIbase: a database of Protein Structural Interactome map (PSIMAP). Sungsam Gong, Giseok Yoon, Insoo Jang Bioinformatics 2005.
- Mapping Protein Family Interactions : Intramolecular and Intermolecular Protein Family Interaction Repertoires in the PDB and Yeast, Jong Park, Michael Lappe & Sarah A. Teichmann,J.M.B (2001).
- Semantic Systems Biology (http://www.semantic-systems-biology.org)

Related fields

Mathematical biology

Mathematical biology is also called **theoretical biology**,[1] and sometimes **biomathematics**. It includes at least four major subfields: *biological mathematical modeling*, *relational biology/complex systems biology (CSB)*, *bioinformatics* and *computational biomodeling/biocomputing*. It is an interdisciplinary academic research field with a wide range of applications in biology, medicine[2] and → biotechnology.[3]

Mathematical biology aims at the mathematical representation, treatment and modeling of biological processes, using a variety of applied mathematical techniques and tools. It has both theoretical and practical applications in biological, biomedical and biotechnology research. For example, in cell biology, protein interactions are often represented as "cartoon" models, which, although easy to visualize, do not accurately describe the systems studied. In order to do this, precise mathematical models are required. By describing the systems in a quantitative manner, their behavior can be better simulated, and hence properties can be predicted that might not be evident to the experimenter.

Importance

Applying mathematics to biology has a long history, but only recently has there been an explosion of interest in the field. Some reasons for this include:

- the explosion of data-rich information sets, due to the → genomics revolution, which are difficult to understand without the use of analytical tools,
- recent development of mathematical tools such as chaos theory to help understand complex, nonlinear mechanisms in biology,
- an increase in computing power which enables calculations and simulations to be performed that were not previously possible, and
- an increasing interest in in silico experimentation due to ethical considerations, risk, unreliability and other complications involved in human and animal research.

For use of basic arithmetics in biology, see relevant topic, such as Serial dilution.

Areas of research

Several areas of specialized research in mathematical and theoretical biology[4] [5] [6] [7] [8] [9] as well as external links to related projects in various universities are concisely presented in the following subsections, including also a large number of appropriate validating references from a list of several thousands of published authors contributing to this field. Many of the included examples are characterised by highly complex, nonlinear, and supercomplex mechanisms, as it is being increasingly recognised that the result of such interactions may only be understood through a combination of mathematical, logical, physical/chemical, molecular and computational models. Due to the wide diversity of specific knowledge involved, biomathematical research is often done in collaboration between mathematicians, biomathematicians, theoretical biologists, physicists,

biophysicists, biochemists, bioengineers, engineers, biologists, physiologists, research physicians, biomedical researchers,oncologists, molecular biologists, geneticists, embryologists, zoologists, chemists, etc.

Computer models and automata theory

A monograph on this topic summarizes an extensive amount of published research in this area up to 1987,[10] including subsections in the following areas: computer modeling in biology and medicine, arterial system models, neuron models, biochemical and oscillation networks, quantum automata, [11]quantum computers in molecular biology and genetics, cancer modelling, neural nets, genetic networks, abstract relational biology, metabolic-replication systems, category theory[12] applications in biology and medicine,[13] automata theory,cellular automata, tessallation models[14] [15] and complete self-reproduction [16], chaotic systems in organisms, relational biology and organismic theories.[17] [18] This published report also includes 390 references to peer-reviewed articles by a large number of authors.[19] [20] [21]

Modeling cell and molecular biology

This area has received a boost due to the growing importance of molecular biology.[22]

- Mechanics of biological tissues[23]
- Theoretical enzymology and enzyme kinetics
- Cancer modelling and simulation [24] [25]
- Modelling the movement of interacting cell populations[26]
- Mathematical modelling of scar tissue formation[27]
- Mathematical modelling of intracellular dynamics[28]
- Mathematical modelling of the cell cycle[29]

Modelling physiological systems

- Modelling of arterial disease [30]
- Multi-scale modelling of the heart [31]

Molecular set theory

Molecular set theory was introduced by Anthony Bartholomay, and its applications were developed in mathematical biology and especially in Mathematical Medicine.[32] Molecular set theory (MST) is a mathematical formulation of the wide-sense chemical kinetics of biomolecular reactions in terms of sets of molecules and their chemical transformations represented by set-theoretical mappings between molecular sets. In a more general sense, MST is the theory of molecular categories defined as categories of molecular sets and their chemical transformations represented as set-theoretical mappings of molecular sets. The theory has also contributed to biostatistics and the formulation of clinical biochemistry problems in mathematical formulations of pathological, biochemical changes of interest to Physiology, Clinical Biochemistry and Medicine.[33] [34]

Population dynamics

Population dynamics has traditionally been the dominant field of mathematical biology. Work in this area dates back to the 19th century. The Lotka–Volterra predator-prey equations are a famous example. In the past 30 years, population dynamics has been complemented by evolutionary game theory, developed first by John Maynard Smith. Under these dynamics, evolutionary biology concepts may take a deterministic mathematical form. Population dynamics overlap with another active area of research in mathematical biology: mathematical epidemiology, the study of infectious disease affecting populations. Various models of viral spread have been proposed and analyzed, and provide important results that may be applied to health policy decisions.

Mathematical methods

A model of a biological system is converted into a system of equations, although the word 'model' is often used synonymously with the system of corresponding equations. The solution of the equations, by either analytical or numerical means, describes how the biological system behaves either over time or at equilibrium. There are many different types of equations and the type of behavior that can occur is dependent on both the model and the equations used. The model often makes assumptions about the system. The equations may also make assumptions about the nature of what may occur.

Mathematical biophysics

The earlier stages of mathematical biology were dominated by mathematical biophysics, described as the application of mathematics in biophysics, often involving specific physical/mathematical models of biosystems and their components or compartments.

The following is a list of mathematical descriptions and their assumptions.

Deterministic processes (dynamical systems)

A fixed mapping between an initial state and a final state. Starting from an initial condition and moving forward in time, a deterministic process will always generate the same trajectory and no two trajectories cross in state space.

- Difference equations – discrete time, continuous state space.
- Ordinary differential equations – continuous time, continuous state space, no spatial derivatives. *See also:* Numerical ordinary differential equations.
- Partial differential equations – continuous time, continuous state space, spatial derivatives. *See also:* Numerical partial differential equations.
- Maps – discrete time, continuous state space.

Stochastic processes (random dynamical systems)

A random mapping between an initial state and a final state, making the state of the system a random variable with a corresponding probability distribution.

- Non-Markovian processes – generalized master equation – continuous time with memory of past events, discrete state space, waiting times of events (or transitions between states) discretely occur and have a generalized probability distribution.
- Jump Markov process – master equation – continuous time with no memory of past events, discrete state space, waiting times between events discretely occur and are exponentially distributed. *See also:* Monte Carlo method for numerical simulation methods, specifically continuous-time Monte Carlo which is also called kinetic Monte

Carlo or the stochastic simulation algorithm.
- Continuous Markov process – stochastic differential equations or a Fokker-Planck equation – continuous time, continuous state space, events occur continuously according to a random Wiener process.

Spatial modelling

One classic work in this area is Alan Turing's paper on morphogenesis entitled *The Chemical Basis of Morphogenesis*, published in 1952 in the Philosophical Transactions of the Royal Society.

- Travelling waves in a wound-healing assay[35]
- Swarming behaviour[36]
- A mechanochemical theory of morphogenesis[37]
- Biological pattern formation[38]
- Spatial distribution modeling using plot samples[39]

Phylogenetics

Phylogenetics is an area of mathematical biology that deals with the reconstruction and analysis of phylogenetic (evolutionary) trees and networks based on inherited characteristics. The main mathematical concepts are trees, X-trees and maximum parsimony trees.

Model example: the cell cycle

The eukaryotic cell cycle is very complex and is one of the most studied topics, since its misregulation leads to cancers. It is possibly a good example of a mathematical model as it deals with simple calculus but gives valid results. Two research groups [40] [41] have produced several models of the cell cycle simulating several organisms. They have recently produced a generic eukaryotic cell cycle model which can represent a particular eukaryote depending on the values of the parameters, demonstrating that the idiosyncrasies of the individual cell cycles are due to different protein concentrations and affinities, while the underlying mechanisms are conserved (Csikasz-Nagy et al., 2006).

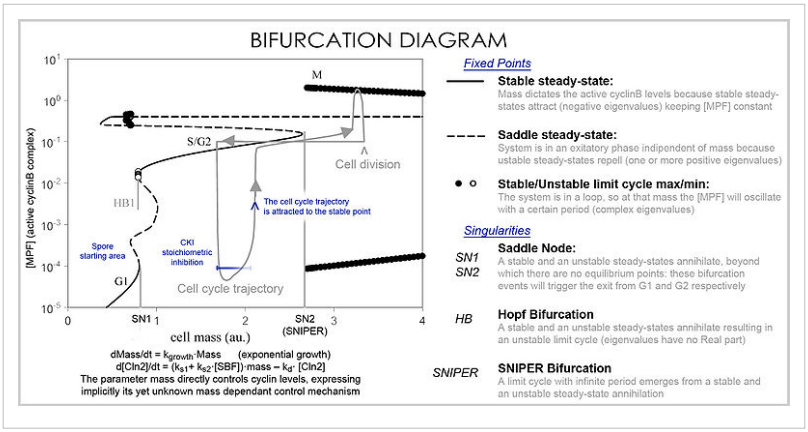
By means of a system of ordinary differential equations these models show the change in time (dynamical system) of the protein inside a single typical cell; this type of model is called a deterministic process (whereas a model describing a statistical distribution of protein concentrations in a population of cells is called a stochastic process).

To obtain these equations an iterative series of steps must be done: first the several models and observations are combined to form a consensus diagram and the appropriate kinetic laws are chosen to write the differential equations, such as rate kinetics for stoichiometric reactions, Michaelis-Menten kinetics for enzyme substrate reactions and Goldbeter–Koshland kinetics for ultrasensitive transcription factors, afterwards the parameters of the equations (rate constants, enzyme efficiency coefficients and Michealis constants) must be fitted to match observations; when they cannot be fitted the kinetic equation is revised and when that is not possible the wiring diagram is modified. The parameters are fitted and validated using observations of both wild type and mutants, such as protein half-life and cell size.

In order to fit the parameters the differential equations need to be studied. This can be done either by simulation or by analysis.

In a simulation, given a starting vector (list of the values of the variables), the progression

of the system is calculated by solving the equations at each time-frame in small increments.



In analysis, the proprieties of the equations are used to investigate the behavior of the system depending of the values of the parameters and variables. A system of differential equations can be represented as a vector field, where each vector described the change (in concentration of two or more protein) determining where and how fast the trajectory (simulation) is heading. Vector fields can have several special points: a stable point, called a sink, that attracts in all directions (forcing the concentrations to be at a certain value), an unstable point, either a source or a saddle point which repels (forcing the concentrations to change away from a certain value), and a limit cycle, a closed trajectory towards which several trajectories spiral towards (making the concentrations oscillate).

A better representation which can handle the large number of variables and parameters is called a bifurcation diagram(Bifurcation theory): the presence of these special steady-state points at certain values of a parameter (e.g. mass) is represented by a point and once the parameter passes a certain value, a qualitative change occurs, called a bifurcation, in which the nature of the space changes, with profound consequences for the protein concentrations: the cell cycle has phases (partially corresponding to G1 and G2) in which mass, via a stable point, controls cyclin levels, and phases (S and M phases) in which the concentrations change independently, but once the phase has changed at a bifurcation event (Cell cycle checkpoint), the system cannot go back to the previous levels since at the current mass the vector field is profoundly different and the mass cannot be reversed back through the bifurcation event, making a checkpoint irreversible. In particular the S and M checkpoints are regulated by means of special bifurcations called a Hopf bifurcation and an infinite period bifurcation.

Mathematical/theoretical biologists

- Pere Alberch
- Anthony F. Bartholomay
- J. T. Bonner
- Jack Cowan
- Gerd B. Müller
- Walter M. Elsasser
- Claus Emmeche
- Andree Ehresmann
- Marc Feldman
- Ronald A. Fisher
- Brian Goodwin
- Bryan Grenfell

- J. B. S. Haldane
- William D. Hamilton
- Lionel G. Harrison
- Michael Hassell
- Sven Erik Jørgensen
- George Karreman
- Stuart Kauffman
- Kalevi Kull
- Herbert D. Landahl
- Richard Lewontin
- Humberto Maturana
- Robert May
- John Maynard Smith
- Howard Pattee
- George R. Price
- Erik Rauch
- Nicolas Rashevsky
- Ronald Brown (mathematician)
- Johannes Reinke
- Robert Rosen
- Rene Thom
- Jakob von Uexküll
- Robert Ulanowicz
- Francisco Varela
- C. H. Waddington
- Arthur Winfree
- Lewis Wolpert
- Sewall Wright
- Christopher Zeeman

Mathematical, theoretical and computational biophysicists

- Nicolas Rashevsky
- Ludwig von Bertalanffy
- Francis Crick
- Manfred Eigen
- Walter Elsasser
- Herbert Frohlich, FRS
- Francois Jacob
- Martin Karplus
- George Karreman
- Herbert D. Landahl
- Ilya, Viscount Prigogine
- SirJohn Randall
- James D. Murray
- Bernard Pullman
- Alberte Pullman
- Erwin Schrodinger

- Klaus Schulten
- Peter Schuster
- Zeno Simon
- D'Arcy Thompson
- Murray Gell-Mann

See also

- Abstract relational biology [42][43] [44]
- Biocybernetics
- → Bioinformatics
- Biologically-inspired computing
- Biostatistics
- Cellular automata[45]
- Coalescent theory
- → Complex systems biology[46] [47] [48]
- Computational biology
- Dynamical systems in biology[49] [50] [51] [52] [53] [54]
- Epidemiology
- Evolution theories and Population Genetics
  - Population genetics models
  - Molecular evolution theories
- Ewens's sampling formula
- Excitable medium
- Mathematical models
  - Molecular modelling
  - Software for molecular modeling
  - Metabolic-replication systems [55][56]
  - Models of Growth and Form
  - Neighbour-sensing model
- Morphometrics
- Organismic systems (OS) [57][58]
- Organismic supercategories [59][60] [61]
- Population dynamics of fisheries
- Protein folding, also blue Gene and folding@home
- Quantum computers
- Quantum genetics
- Relational biology [62]
- → Self-reproduction[63] (also called self-replication in a more general context).
- Computational gene models
- → Systems biology[64]
- Theoretical biology[65]
- Topological models of morphogenesis
  - DNA topology
  - DNA sequencing theory

    For use of basic arithmetics in biology, see relevant topic, such as Serial dilution.
- Biographies

- Charles Darwin
- D'Arcy Thompson
- Joseph Fourier
- Charles S. Peskin
- Nicolas Rashevsky [66]
- Robert Rosen
- Rosalind Franklin
- Francis Crick
- René Thom
- Vito Volterra

References

- Nicolas Rashevsky. (1938)., *Mathematical Biophysics*. Chicago: University of Chicago Press.
- Robert Rosen, Dynamical system theory in biology. New York, Wiley-Interscience (1970) ISBN 0471735507 [67]
- Israel, G., 2005, "Book on mathematical biology" in Grattan-Guinness, I., ed., *Landmark Writings in Western Mathematics*. Elsevier: 936-44.
- Israel, G (1988), "On the contribution of Volterra and Lotka to the development of modern biomathematics. [68]", *History and philosophy of the life sciences* **10** (1): 37-49, PMID:3045853, http://www.ncbi.nlm.nih.gov/pubmed/3045853
- Scudo, F M (1971), "Vito Volterra and theoretical ecology. [69]", *Theoretical population biology* **2** (1): 1-23, 1971 Mar, PMID:4950157, http://www.ncbi.nlm.nih.gov/pubmed/4950157
- S.H. Strogatz, *Nonlinear dynamics and Chaos: Applications to Physics, Biology, Chemistry, and Engineering.* Perseus, 2001, ISBN 0-7382-0453-6
- N.G. van Kampen, *Stochastic Processes in Physics and Chemistry*, North Holland., 3rd ed. 2001, ISBN 0-444-89349-0
- I. C. Baianu., Computer Models and Automata Theory in Biology and Medicine., *Monograph*, Ch.11 in M. Witten (Editor), *Mathematical Models in Medicine*, vol. **7**., Vol. **7**: 1513-1577 (1987),Pergamon Press:New York, (updated by Hsiao Chen Lin in 2004[70] ,[71] ,[72] ISBN 0080363776 [73] .
- P.G. Drazin, *Nonlinear systems*. C.U.P., 1992. ISBN 0-521-40668-4
- L. Edelstein-Keshet, *Mathematical Models in Biology*. SIAM, 2004. ISBN 0-07-554950-6
- G. Forgacs and S. A. Newman, *Biological Physics of the Developing Embryo*. C.U.P., 2005. ISBN 0-521-78337-2
- A. Goldbeter, *Biochemical oscillations and cellular rhythms*. C.U.P., 1996. ISBN 0-521-59946-6
- L.G. Harrison, *Kinetic theory of living pattern*. C.U.P., 1993. ISBN 0-521-30691-4
- F. Hoppensteadt, *Mathematical theories of populations: demographics, genetics and epidemics*. SIAM, Philadelphia, 1975 (reprinted 1993). ISBN 0-89871-017-0
- D.W. Jordan and P. Smith, *Nonlinear ordinary differential equations*, 2nd ed. O.U.P., 1987. ISBN 0-19-856562-3
- J.D. Murray, *Mathematical Biology*. Springer-Verlag, 3rd ed. in 2 vols.: *Mathematical Biology: I. An Introduction*, 2002 ISBN 0-387-95223-3; *Mathematical Biology: II. Spatial Models and Biomedical Applications*, 2003 ISBN 0-387-95228-4.

- E. Renshaw, *Modelling biological populations in space and time*. C.U.P., 1991. ISBN 0-521-44855-7
- S.I. Rubinow, *Introduction to mathematical biology*. John Wiley, 1975. ISBN 0-471-74446-8
- L.A. Segel, *Modeling dynamic phenomena in molecular and cellular biology*. C.U.P., 1984. ISBN 0-521-27477-X
- L. Preziosi, *Cancer Modelling and Simulation*. Chapman Hall/CRC Press, 2003. ISBN 1-58488-361-8.

Lists of references

- A general list of Theoretical biology/Mathematical biology references, including an updated list of actively contributing authors[74] .
- A list of references for applications of category theory in relational biology[75] .
- An updated list of publications of theoretical biologist Robert Rosen[76]

External

- F. Hoppensteadt, *Getting Started in Mathematical Biology* [77]. Notices of American Mathematical Society, Sept. 1995.
- M. C. Reed, *Why Is Mathematical Biology So Hard?* [78] Notices of American Mathematical Society, March, 2004.
- R. M. May, *Uses and Abuses of Mathematics in Biology* [79]. Science, February 6, 2004.
- J. D. Murray, *How the leopard gets its spots?* [80] Scientific American, 258(3): 80-87, 1988.
- S. Schnell, R. Grima, P. K. Maini, *Multiscale Modeling in Biology* [81], American Scientist, Vol 95, pages 134-142, March-April 2007.
- Chen KC et al. Integrative analysis of cell cycle control in budding yeast. Mol Biol Cell. 2004 Aug;15(8):3841-62.
- Csikász-Nagy A et al. Analysis of a generic model of eukaryotic cell-cycle regulation. Biophys J. 2006 Jun 15;90(12):4361-79.
- Fuss H, et al. Mathematical models of cell cycle regulation. Brief Bioinform. 2005 Jun;6(2):163-77.
- Lovrics A et al. *Time scale and dimension analysis of a budding yeast cell cycle model.* [82] BMC Bioinform. 2006 Nov 9;7:494.

Notes: Inline and online

[1] Mathematical Biology and Theoretical Biophysics-An Outline: What is Life? http://planetmath.org/?op=getobj&from=objects&id=10921
[2] http://www.kli.ac.at/theorylab/EditedVol/W/WittenM1987a.html
[3] http://en.scientificcommons.org/1857372
[4] http://www.kli.ac.at/theorylab/index.html
[5] http://www.springerlink.com/content/w2733h7280521632/
[6] http://en.scientificcommons.org/1857371
[7] http://cogprints.org/3687/
[8] "Research in Mathematical Biology" (http://www.maths.gla.ac.uk/research/groups/biology/kal.htm). Maths.gla.ac.uk. . Retrieved on 2008-09-10.
[9] http://acube.org/volume_23/v23-1p11-36.pdf J. R. Junck. Ten Equations that Changed Biology: Mathematics in Problem-Solving Biology Curricula, *Bioscene*, (1997), 1-36
[10] http://en.scientificcommons.org/1857371
[11] http://planetphysics.org/encyclopedia/QuantumAutomaton.html

[12] http://planetphysics.org/encyclopedia/BibliographyForCategoryTheoryAndAlgebraicTopologyApplicationsInTheoreticalPhysics.html
[13] http://planetphysics.org/encyclopedia/BibliographyForMathematicalBiophysicsAndMathematicalMedicine.html
[14] *Modern Cellular Automata* by Kendall Preston and M. J. B. Duff http://books.google.co.uk/books?id=l0_0q_e-u_UC&dq=cellular+automata+and+tessalation&pg=PP1&ots=ciXYCF3AYm&source=citation&sig=CtaUDhisM7MalS7rZfXvp689y-8&hl=en&sa=X&oi=book_result&resnum=12&ct=result
[15] http://mathworld.wolfram.com/DualTessellation.html
[16] http://planetphysics.org/encyclopedia/ETACAxioms.html
[17] Baianu, I. C. 1987, Computer Models and Automata Theory in Biology and Medicine., in M. Witten (ed.),*Mathematical Models in Medicine*, vol. **7**., Ch.11 Pergamon Press, New York, 1513-1577. http://cogprints.org/3687/
[18] http://www.kli.ac.at/theorylab/EditedVol/W/WittenM1987a.html
[19] http://www.springerlink.com/content/w2733h7280521632/
[20] Currently available for download as an updated PDF: http://cogprints.ecs.soton.ac.uk/archive/00003718/01/COMPUTER_SIMULATIONCOMPUTABILITYBIOSYSTEMSrefnew.pdf
[21] http://planetphysics.org/encyclopedia/BibliographyForMathematicalBiophysics.html
[22] "Research in Mathematical Biology" (http://www.maths.gla.ac.uk/research/groups/biology/kal.htm). Maths.gla.ac.uk. . Retrieved on 2008-09-10.
[23] http://www.maths.gla.ac.uk/~rwo/research_areas.htm
[24] http://www.springerlink.com/content/71958358k273622q/
[25] http://calvino.polito.it/~mcrtn/
[26] http://www.ma.hw.ac.uk/~jas/researchinterests/index.html
[27] http://www.ma.hw.ac.uk/~jas/researchinterests/scartissueformation.html
[28] http://www.sbi.uni-rostock.de/dokumente/p_gilles_paper.pdf
[29] http://mpf.biol.vt.edu/Research.html
[30] http://www.maths.gla.ac.uk/~nah/research_interests.html
[31] http://www.integrativebiology.ox.ac.uk/heartmodel.html
[32] http://planetphysics.org/encyclopedia/CategoryOfMolecularSets2.html
[33] Representation of Uni-molecular and Multimolecular Biochemical Reactions in terms of Molecular Set Transformations http://planetmath.org/?op=getobj&from=objects&id=10770
[34] http://planetphysics.org/encyclopedia/CategoryOfMolecularSets2.html
[35] http://www.maths.ox.ac.uk/~maini/public/gallery/twwha.htm
[36] http://www.math.ubc.ca/people/faculty/keshet/research.html
[37] http://www.maths.ox.ac.uk/~maini/public/gallery/mctom.htm
[38] http://www.maths.ox.ac.uk/~maini/public/gallery/bpf.htm
[39] http://links.jstor.org/sici?sici=0030-1299%28199008%2958%3A3%3C257%3ASDOTMU%3E2.0.CO%3B2-S&size=LARGE&origin=JSTOR-enlargePage
[40] "The JJ Tyson Lab" (http://mpf.biol.vt.edu/Tyson Lab.html). Virginia Tech. . Retrieved on 2008-09-10.
[41] "The Molecular Network Dynamics Research Group" (http://cellcycle.mkt.bme.hu/). Budapest University of Technology and Economics. .
[42] http://www.kli.ac.at/theorylab/ALists/Authors_R.html
[43] http://planetphysics.org/encyclopedia/AbstractRelationalBiologyARB.html
[44] http://www.kli.ac.at/theorylab/EditedVol/M/MatsunoKDose_84.html
[45] Baianu, I. C. 1987, Computer Models and Automata Theory in Biology and Medicine., in M. Witten (ed.),*Mathematical Models in Medicine*, vol. **7**., Ch.11 Pergamon Press, New York, 1513-1577. http://www.springerlink.com/content/w2733h7280521632/
[46] http://www.springerlink.com/content/v1rt05876h74v607/?p=2bd3993c33644512ba7069ed7fad0046&pi=1
[47] http://www.springerlink.com/content/j7t56r530140r88p/?p=2bd3993c33644512ba7069ed7fad0046&pi=3
[48] http://www.springerlink.com/content/98303486x3l07jx3/
[49] Robert Rosen, *Dynamical system theory in biology.* New York, Wiley-Interscience (1970) ISBN 0471735507 http://www.worldcat.org/oclc/101642
[50] http://www.springerlink.com/content/j7t56r530140r88p/?p=2bd3993c33644512ba7069ed7fad0046&pi=3
[51] http://cogprints.org/3674/
[52] http://cogprints.org/3829/
[53] http://www.ncbi.nlm.nih.gov/pubmed/4327361
[54] http://www.springerlink.com/content/98303486x3l07jx3/
[55] http://planetphysics.org/encyclopedia/RSystemsCategoryOfM.html

[56] http://www.kli.ac.at/theorylab/ALists/Authors_R.html
[57] http://planetphysics.org/encyclopedia/OrganismicSetTheory.html
[58] Organisms as Super-complex Systems http://planetmath.org/?op=getobj&from=objects&id=10890
[59] http://planetphysics.org/encyclopedia/OrganismicSetTheory.html
[60] http://www.springerlink.com/content/98303486x3l07jx3/
[61] http://planetmath.org/encyclopedia/SupercategoriesOfComplexSystems.html
[62] http://planetmath.org/?op=getobj&from=objects&id=10921
[63] http://planetmath.org/?method=l2h&from=objects&
name=NaturalTransformationsOfOrganismicStructures&op=getobj
[64] http://www.kli.ac.at/theorylab/ALists/Authors_R.html
[65] http://www.kli.ac.at/theorylab/index.html
[66] http://planetphysics.org/encyclopedia/NicolasRashevsky.html
[67] http://www.worldcat.org/oclc/101642
[68] http://www.ncbi.nlm.nih.gov/pubmed/3045853
[69] http://www.ncbi.nlm.nih.gov/pubmed/4950157
[70] http://cogprints.org/3718/1/COMPUTER_SIMULATIONCOMPUTABILITYBIOSYSTEMSrefnew.pdf
[71] http://www.springerlink.com/content/w2733h7280521632/
[72] http://www.springerlink.com/content/n8gw445012267381/
[73] http://www.bookfinder.com/dir/i/Mathematical_Models_in_Medicine/0080363776/
[74] http://www.kli.ac.at/theorylab/index.html
[75] http://planetmath.org/?method=l2h&from=objects&id=10746&op=getobj
[76] Publications list for Robert Rosen http://www.people.vcu.edu/~mikuleck/rosen.htm
[77] http://www.ams.org/notices/199509/hoppensteadt.pdf
[78] http://www.resnet.wm.edu/~jxshix/math490/reed.pdf
[79] http://www.resnet.wm.edu/~jxshix/math490/may.pdf
[80] http://www.resnet.wm.edu/~jxshix/math490/murray.doc
[81] http://eprints.maths.ox.ac.uk/567/01/224.pdf
[82] http://www.biomedcentral.com/content/pdf/1471-2105-7-494.pdf

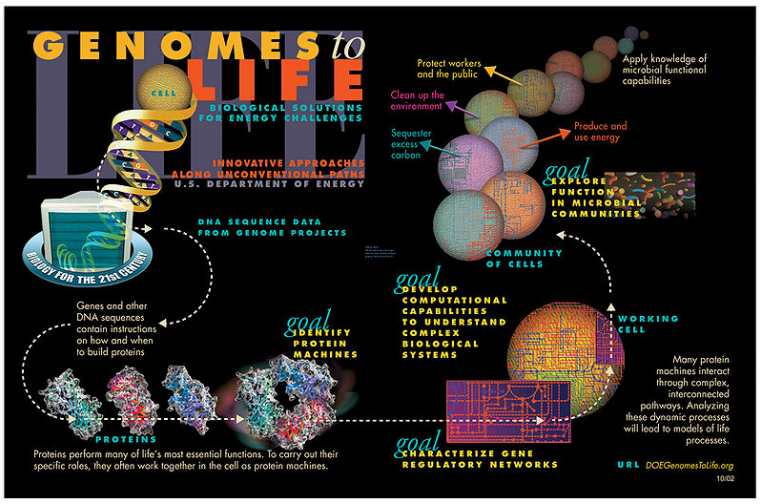
External links

- Theoretical and mathematical biology website (http://www.kli.ac.at/theorylab/index.html)
- Complexity Discussion Group (http://www.complex.vcu.edu/)
- Integrative cancer biology modeling and Complex systems biology (http://fs512.fshn.uiuc.edu/ComplexSystemsBiology.htm)
- UCLA Biocybernetics Laboratory (http://biocyb.cs.ucla.edu/research.html)
- TUCS Computational Biomodelling Laboratory (http://www.tucs.fi/research/labs/combio.php)
- Nagoya University Division of Biomodeling (http://www.agr.nagoya-u.ac.jp/english/e3senko-1.html)
- Technische Universiteit Biomodeling and Informatics (http://www.bmi2.bmt.tue.nl/Biomedinf/)
- BioCybernetics Wiki, a vertical wiki on biomedical cybernetics and systems biology (http://wiki.biological-cybernetics.de)
- Society for Mathematical Biology (http://www.smb.org/)
- Bulletin of Mathematical Biology (http://www.springerlink.com/content/119979/)
- European Society for Mathematical and Theoretical Biology (http://www.esmtb.org/)
- Journal of Mathematical Biology (http://www.springerlink.com/content/100436/)
- Biomathematics Research Centre at University of Canterbury (http://www.math.canterbury.ac.nz/bio/)
- Centre for Mathematical Biology at Oxford University (http://www.maths.ox.ac.uk/cmb/)

- Mathematical Biology at the National Institute for Medical Research (http://mathbio.nimr.mrc.ac.uk/)
- Institute for Medical BioMathematics (http://www.imbm.org/)
- *Mathematical Biology Systems of Differential Equations* (http://eqworld.ipmnet.ru/en/solutions/syspde/spde-toc2.pdf) from EqWorld: The World of Mathematical Equations
- Systems Biology Workbench - a set of tools for modelling biochemical networks (http://sbw.kgi.edu)
- The Collection of Biostatistics Research Archive (http://www.biostatsresearch.com/repository/)
- Statistical Applications in Genetics and Molecular Biology (http://www.bepress.com/sagmb/)
- The International Journal of Biostatistics (http://www.bepress.com/ijb/)

Systems biology

Systems biology is a biology-based inter-disciplinary study field that focuses on the systematic study of complex interactions in biological systems, thus using a new perspective (holism instead of reduction) to study them. Particularly from year 2000 onwards, the term is used widely in the biosciences, and in a variety of contexts. Because the scientific method has been used primarily toward reductionism, one of the goals of systems biology is to discover new emergent properties that may arise from the systemic view used by this discipline in order to understand better the entirety of processes that happen in a biological system.



Example of systems biology research.

Overview

Systems biology can be considered from a number of different aspects:

- Some sources discuss systems biology as a **field of study**, particularly, the study of the interactions between the components of *biological systems*, and how these interactions give rise to the function and behavior of that system (for example, the enzymes and metabolites in a → metabolic pathway).[1] [2]
- Other sources consider systems biology as a **paradigm**, usually defined in antithesis to the so-called reductionist paradigm, although fully consistent with the scientific method. The distinction between the two paradigms is referred to in these quotations:

  "*The reductionist approach has successfully identified most of the components and many of the interactions but, unfortunately, offers no convincing concepts or methods*

to understand how system properties emerge...the pluralism of causes and effects in biological networks is better addressed by observing, through quantitative measures, multiple components simultaneously and by rigorous data integration with mathematical models" Science[3]

"*Systems biology...is about putting together rather than taking apart, integration rather than reduction. It requires that we develop ways of thinking about integration that are as rigorous as our reductionist programmes, but different....It means changing our philosophy, in the full sense of the term*" Denis Noble[4]

- Still other sources view systems biology in terms of the **operational protocols used for performing research**, namely a cycle composed of theory, analytic or computational modelling to propose specific testable hypotheses about a biological system, experimental validation, and then using the newly acquired quantitative description of cells or cell processes to refine the computational model or theory.[5] [6] Since the objective is a model of the interactions in a system, the experimental techniques that most suit systems biology are those that are system-wide and attempt to be as complete as possible. Therefore, transcriptomics, metabolomics, → proteomics and high-throughput techniques are used to collect quantitative data for the construction and validation of models.
- Engineers consider systems biology as the application of dynamical systems theory to molecular biology.
- Finally, some sources see it as a **socioscientific phenomenon** defined by the strategy of pursuing integration of complex data about the interactions in biological systems from diverse experimental sources using interdisciplinary tools and personnel.

This variety of viewpoints is illustrative of the fact that systems biology refers to a cluster of peripherally overlapping concepts rather than a single well-delineated field. However the term has widespread currency and popularity as of 2007, with chairs and institutes of systems biology proliferating worldwide (Such as the Institute for Systems Biology).

History

Systems biology finds its roots in:

- the quantitative modelling of enzyme kinetics, a discipline that flourished between 1900 and 1970,
- the simulations developed to study neurophysiology, and
- control theory and cybernetics.

One of the theorists who can be seen as a precursor of systems biology is Ludwig von Bertalanffy with his general systems theory. One of the first numerical simulations in biology was published in 1952 by the British neurophysiologists and Nobel prize winners Alan Lloyd Hodgkin and Andrew Fielding Huxley, who constructed a mathematical model that explained the action potential propagating along the axon of a neuronal cell.[7] Their model described a cellular function emerging from the interaction between two different molecular components, a potassium and a sodium channels, and can therefore be seen as the beginning of computational systems biology.[8] In 1960, Denis Noble developed the first computer model of the heart pacemaker.[9]

The formal study of systems biology, as a distinct discipline, was launched by systems theorist Mihajlo Mesarovic in 1966 with an international symposium at the Case Institute of

Technology in Cleveland, Ohio entitled "Systems Theory and Biology."[10] [11]

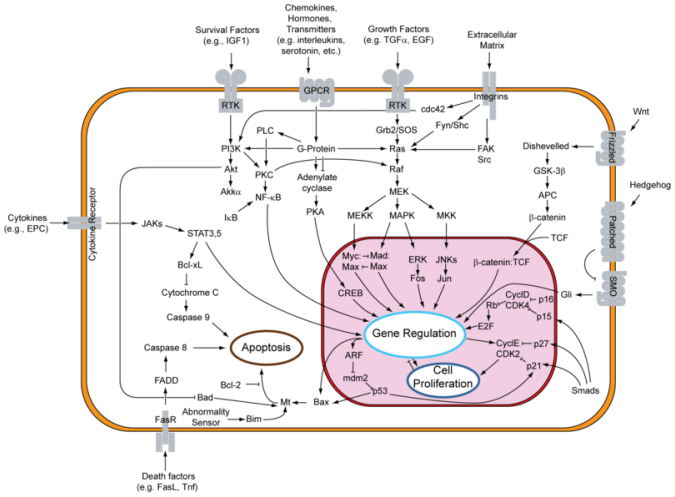
The 1960s and 1970s saw the development of several approaches to study complex molecular systems, such as the Metabolic Control Analysis and the biochemical systems theory. The successes of molecular biology throughout the 1980s, coupled with a skepticism toward theoretical biology, that then promised more than it achieved, caused the quantitative modelling of biological processes to become a somewhat minor field.

However the birth of functional genomics in the 1990s meant that large quantities of high quality data became available, while the computing power exploded, making more realistic models possible. In 1997, the group of Masaru Tomita published the first quantitative model of the metabolism of a whole (hypothetical) cell.

Around the year 2000, when Institutes of Systems Biology were established in Seattle and Tokyo, systems biology emerged as a movement in its own right, spurred on by the completion of various genome projects, the large increase in data from the omics (e.g. → genomics and → proteomics) and the accompanying advances in high-throughput experiments and → bioinformatics. Since then, various research institutes dedicated to systems biology have been developed. As of summer 2006, due to a shortage of people in systems biology[12] several doctoral training centres in systems biology have been established in many parts of the world.

Techniques associated with systems biology

According to the interpretation of System Biology as the ability to obtain, integrate and analyze complex data from multiple experimental sources using interdisciplinary tools, some typical technology platforms are:



Overview of signal transduction pathways

- Transcriptomics: whole cell or tissue gene expression measurements by DNA microarrays or serial analysis of gene expression
- → Proteomics: complete identification of proteins and protein expression patterns of a cell or tissue through two-dimensional gel electrophoresis and mass spectrometry or multi-dimensional protein identification techniques (advanced HPLC systems coupled with mass spectrometry). Sub disciplines include phosphoproteomics, glycoproteomics and other methods to detect chemically modified proteins.
- Metabolomics: identification and measurement of all small-molecules metabolites within a cell or tissue
- Glycomics: identification of the entirety of all carbohydrates in a cell or tissue.

In addition to the identification and quantification of the above given molecules further techniques analyze the dynamics and interactions within a cell. This includes:

- → Interactomics which is used mostly in the context of protein-protein interaction but in theory encompasses interactions between all molecules within a cell,
- Fluxomics, which deals with the dynamic changes of molecules within a cell over time,
- Biomics: systems analysis of the biome.

The investigations are frequently combined with large scale perturbation methods, including gene-based (RNAi, mis-expression of wild type and mutant genes) and chemical approaches using small molecule libraries. Robots and automated sensors enable such large-scale experimentation and data acquisition. These technologies are still emerging and many face problems that the larger the quantity of data produced, the lower the quality. A wide variety of quantitative scientists (computational biologists, statisticians, mathematicians, computer scientists, engineers, and physicists) are working to improve the quality of these approaches and to create, refine, and retest the models to accurately reflect observations.

The investigations of a single level of biological organization (such as those listed above) are usually referred to as Systematic Systems Biology. Other areas of Systems Biology includes Integrative Systems Biology, which seeks to integrate different types of information to advance the understanding the biological whole, and Dynamic Systems Biology, which aims to uncover how the biological whole changes over time (during evolution, for example, the onset of disease or in response to a perturbation). Functional Genomics may also be considered a sub-field of Systems Biology.

The systems biology approach often involves the development of mechanistic models, such as the reconstruction of dynamic systems from the quantitative properties of their elementary building blocks.[13] [14] For instance, a cellular network can be modelled mathematically using methods coming from chemical kinetics and control theory. Due to the large number of parameters, variables and constraints in cellular networks, numerical and computational techniques are often used. Other aspects of computer science and informatics are also used in systems biology. These include new forms of computational model, such as the use of process calculi to model biological processes, the integration of information from the literature, using techniques of information extraction and text mining, the development of online databases and repositories for sharing data and models (such as BioModels Database), approaches to database integration and software interoperability via loose coupling of software, websites and databases[15] and the development of syntactically and semantically sound ways of representing biological models, such as the Systems Biology Markup Language (SBML).

See also

Related fields

- → Complex systems biology
- Complex systems
- Complex systems biology
- → Bioinformatics
- Biological network inference
- Biological systems engineering
- Biomedical cybernetics
- Biostatistics
- Theoretical Biophysics
- Relational Biology
- Translational Research
- Computational biology
- Computational systems biology
- Scotobiology
- Synthetic biology
- Systems biology modeling
- Systems ecology
- Systems immunology

Related terms

- Life
- Artificial life
- Gene regulatory network
- → Metabolic network modelling
- Living systems theory
- Network Theory of Aging
- Regulome
- Systems Biology Markup Language (SBML)
- SBO
- Viable System Model
- Antireductionism

Systems biologists

- Category:Systems biologists

Lists

- Category:Systems biologists
- List of systems biology conferences
- List of omics topics in biology
- List of publications in systems biology
- List of systems biology research groups

References

[1] Snoep J.L. and Westerhoff H.V.; Alberghina L. and Westerhoff H.V. (Eds.) (2005.). "From isolation to integration, a systems biology approach for building the Silicon Cell". *Systems Biology: Definitions and Perspectives*: p7, Springer-Verlag.

[2] "Systems Biology - the 21st Century Science" (http://www.systemsbiology.org/Intro_to_ISB_and_Systems_Biology/Systems_Biology_--_the_21st_Century_Science). .

[3] Sauer, U. et al. (27 April 2007). "Getting Closer to the Whole Picture". *Science* **316**: 550. doi: 10.1126/science.1142502 (http://dx.doi.org/10.1126/science.1142502). PMID 17463274.

[4] Denis Noble (2006). *The Music of Life: Biology beyond the genome*. Oxford University Press. ISBN 978-0199295739. p21

[5] "Systems Biology: Modelling, Simulation and Experimental Validation" (http://www.bbsrc.ac.uk/science/areas/ebs/themes/main_sysbio.html). .

[6] Kholodenko B.N., Bruggeman F.J., Sauro H.M.; Alberghina L. and Westerhoff H.V.(Eds.) (2005.). "Mechanistic and modular approaches to modeling and inference of cellular regulatory networks". *Systems Biology: Definitions and Perspectives*: p143, Springer-Verlag.

[7] Hodgkin AL, Huxley AF (1952). "A quantitative description of membrane current and its application to conduction and excitation in nerve". *J Physiol* **117**: 500–544. PMID 12991237.

[8] Le Novere (2007). "The long journey to a Systems Biology of neuronal function". *BMC Systems Biology* **1**: 28. doi: 10.1186/1752-0509-1-28 (http://dx.doi.org/10.1186/1752-0509-1-28).

[9] Noble D (1960). "Cardiac action and pacemaker potentials based on the Hodgkin-Huxley equations". *Nature* **188**: 495–497. doi: 10.1038/188495b0 (http://dx.doi.org/10.1038/188495b0). PMID 13729365.

[10] Mesarovic, M. D. (1968). *Systems Theory and Biology*. Springer-Verlag.

[11] " A Means Toward a New Holism (http://www.jstor.org/view/00368075/ap004022/00a00220/0)". *Science* **161** (3836): 34-35. doi: 10.1126/science.161.3836.34 (http://dx.doi.org/10.1126/science.161.3836.34). .

[12] "Working the Systems" (http://sciencecareers.sciencemag.org/career_development/previous_issues/articles/2006_03_03/working_the_systems/(parent)/158). .

[13] Gardner, TS; di Bernardo D, Lorenz D and Collins JJ (4 July 2003). "Inferring genetic networks and identifying compound of action via expression profiling". *Science* **301**: 102-1005. doi: 10.1126/science.1081900 (http://dx.doi.org/10.1126/science.1081900). PMID 12843395.

[14] di Bernardo, D; Thompson MJ, Gardner TS, Chobot SE, Eastwood EL, Wojtovich AP, Elliot SJ, Schaus SE and Collins JJ (March 2005). "Chemogenomic profiling on a genome-wide scale using reverse-engineered gene

networks". *Nature Biotechnology* **23**: 377-383. doi: 10.1038/nbt1075 (http://dx.doi.org/10.1038/nbt1075). PMID 15765094.

[15] such as Gaggle (http://gaggle.systemsbiology.net), SBW (http://sys-bio.org)), or commercial suits, e.g., MetaCore (http://www.genego.com/metacore.php) and MetaDrug (http://www.genego.com/metadrug.php)

Further reading

Books

- Hiroaki Kitano (editor). *Foundations of Systems Biology.* MIT Press: 2001. ISBN 0-262-11266-3
- CP Fall, E Marland, J Wagner and JJ Tyson (Editors). "Computational Cell Biology." Springer Verlag: 2002 ISBN 0-387-95369-8
- G Bock and JA Goode (eds).*In Silico*" *Simulation of Biological Processes*, Novartis Foundation Symposium 247. John Wiley & Sons: 2002. ISBN 0-470-84480-9
- E Klipp, R Herwig, A Kowald, C Wierling, and H Lehrach. *Systems Biology in Practice.* Wiley-VCH: 2005. ISBN 3-527-31078-9
- L. Alberghina and H. Westerhoff (Editors) – *Systems Biology: Definitions and Perspectives*, Topics in Current Genetics 13, Springer Verlag (2005), ISBN 13: 978-3540229681
- A Kriete, R Eils. *Computational Systems Biology.*, Elsevier - Academic Press: 2005. ISBN 0-12-088786-X
- K. Sneppen and G. Zocchi, (2005) *Physics in Molecular Biology*, Cambridge University Press, ISBN 0-521-84419-3
- D. Noble, *The Music of life. Biology beyond the genome* Oxford University Press (http://www.musicoflife.co.uk/) 2006. ISBN 0199295735, ISBN 978-0199295739
- Z. Szallasi, J. Stelling, and V.Periwal (eds.) System Modeling in Cellular Biology: From Concepts to Nuts and Bolts (Hardcover), MIT Press: 2006, ISBN 0-262-19548-8
- B Palsson, *Systems Biology - Properties of Reconstructed Networks.* Cambridge University Press: 2006. (http://gcrg.ucsd.edu/book/index.html) ISBN 978-0-521-85903-5
- K Kaneko. *Life: An Introduction to Complex Systems Biology.* Springer: 2006. ISBN 3540326669
- U Alon. *An Introduction to Systems Biology: Design Principles of Biological Circuits.* CRC Press: 2006. ISBN 1-58488-642-0 - emphasis on Network Biology (For a comparative review of Alon, Kaneko and Palsson see Werner, E. (March 29, 2007). " All systems go (http://www.nature.com/nature/journal/v446/n7135/pdf/446493a.pdf)" (PDF). *Nature* **446**: 493-494. doi: 10.1038/446493a (http://dx.doi.org/10.1038/446493a). http://www.nature.com/nature/journal/v446/n7135/pdf/446493a.pdf.)
- Andriani Daskalaki (editor) "Handbook of Research on Systems Biology Applications in Medicine" Medical Information Science Reference, October 2008 ISBN : 978-1-60566-076-9

Journals

- BMC Systems Biology (http://www.biomedcentral.com/bmcsystbiol) - open access journal on systems biology
- Molecular Systems Biology (http://www.nature.com/msb) - open access journal on systems biology
- IET Systems Biology (http://www.ietdl.org/IET-SYB) - not open access journal on systems biology

Articles

- Binnewies, Tim Terence, Miller, WG, Wang, G. *The complete genome sequence and analysis of the human pathogen Campylobacter lari* (http://www.bio.dtu.dk/English/Publications/1/all.aspx?lg=showcommon&id=231324). Published in journal: Foodborne Pathog Disease (ISSN 1535-3141) , vol: 5, issue: 4, pages: 371-386, 2008, Mary Ann Liebert, Inc. Publishers.
- M. Tomita, Hashimoto K, Takahashi K, Shimizu T, Matsuzaki Y, Miyoshi F, Saito K, Tanida S, Yugi K, Venter JC, Hutchison CA. E-CELL: Software Environment for Whole Cell Simulation. Genome Inform Ser Workshop Genome Inform. 1997;8:147-155. (http://web.sfc.keio.ac.jp/~mt/mt-lab/publications/Paper/ecell/bioinfo99/btc007_gml.html)
- ScienceMag.org (http://www.sciencemag.org/content/vol295/issue5560/) - Special Issue: Systems Biology, *Science*, Vol 295, No 5560, March 1, 2002
- Marc Vidal and Eileen E. M. Furlong. Nature Reviews Genetics 2004 From OMICS to systems biology (http://www.nature.com/nrg/journal/v5/n10/poster/omics/index.html)
- Marc Facciotti, Richard Bonneau, Leroy Hood and Nitin Baliga. Current Genomics 2004 Systems Biology Experimental Design - Considerations for Building Predictive Gene Regulatory Network Models for Prokaryotic Systems (http://www.ingentaconnect.com/content/ben/cg/2004/00000005/00000007/art00002)
- Katia Basso, Adam A Margolin, Gustavo Stolovitzky, Ulf Klein, Riccardo Dalla-Favera, Andrea Califano, (2005) "Reverse engineering of regulatory networks in human B cells" (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=15778709&query_hl=7). Nat Genet;37(4):382-90
- Mario Jardon Systems Biology: An Overview (http://www.scq.ubc.ca/?p=253) - a review from the Science Creative Quarterly, 2005
- Johnjoe McFadden, Guardian.co.uk (http://www.guardian.co.uk/life/science/story/0,12996,1477776,00.html) - 'The unselfish gene: The new biology is reasserting the primacy of the whole organism - the individual - over the behaviour of isolated genes', *The Guardian* (May 6, 2005)
- Pharoah, M.C. (online). Looking to systems theory for a reductive explanation of phenomenal experience and evolutionary foundations for higher order thought (http://homepage.ntlworld.com/m.pharoah/) Retrieved Jan, 15 2008.
- WTEC Panel Report on International Research and Development in Systems Biology (http://www.wtec.org/sysbio/welcome.htm) (2005)
- E. Werner, "The Future and Limits of Systems Biology", Science STKE (http://stke.sciencemag.org/content/vol2005/issue278/) 2005, pe16 (2005).
- Francis J. Doyle and Jörg Stelling, "Systems interface biology" (http://www.journals.royalsoc.ac.uk/openurl.asp?genre=article&doi=10.1098/rsif.2006.0143) *J. R. Soc. Interface* Vol 3, No 10 2006

- Kahlem, P. and Birney E. (2006). "Dry work in a wet world: computation in systems biology." Mol Syst Biol 2: 40. (http://www.nature.com/doifinder/10.1038/msb4100080)
- E. Werner, "All systems go" (http://www.nature.com/nature/journal/v446/n7135/pdf/446493a.pdf), "Nature" (http://www.nature.com/nature/journal/v446/n7135/index.html) vol 446, pp 493–494, March 29, 2007. (Review of three books (Alon, Kaneko, and Palsson) on systems biology.)
- Santiago Schnell, Ramon Grima, Philip K. Maini, "Multiscale Modeling in Biology" (http://www.americanscientist.org/template/AssetDetail/assetid/54784), American Scientist, Vol 95, pages 134-142, March-April 2007.
- TS Gardner, D di Bernardo, D Lorenz and JJ Collins. "Inferring genetic networks and identifying compound of action via expression profiling." (http://www.bu.edu/abl/publications.html) Science 301: 102-105 (2003).
- Jeffery C. Way and Pamela A. Silver, Why We Need Systems Biology (http://cs.calstatela.edu/wiki/images/9/9b/Silver.pdf)
- H.S. Wiley, "Systems Biology - Beyond the Buzz." The Scientist (http://www.the-scientist.com/2006/6/1/52/1/). June 2006.]
- Nina Flanagan, "Systems Biology Alters Drug Development." (http://www.genengnews.com/articles/chitem.aspx?aid=2337) Genetic Engineering & Biotechnology News, January 2008

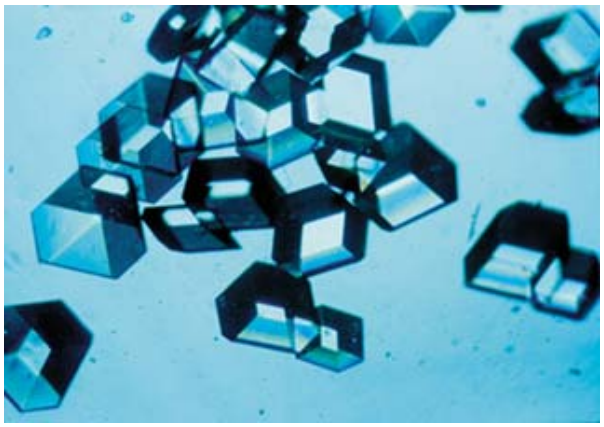
External links

- Systems Biology - BioChemWeb.org (http://www.biochemweb.org/systems.shtml)
- Systems Biology Portal (http://www.systems-biology.org/) - administered by the Systems Biology Institute
- Semantic Systems Biology (http://www.semantic-systems-biology.org)
- SystemsX.ch (http://www.systemsx.ch/) - The Swiss Initiative in Systems Biology
- Systems Biology at the Pacific Northwest National Laboratory (http://www.sysbio.org/)

Biotechnology

Biotechnology is technology based on biology, especially when used in agriculture, food science, and medicine. United Nations Convention on Biological Diversity defines biotechnology as:[1]

> Any technological application that uses biological systems, dead organisms, or derivatives thereof, to make or modify products or processes for specific use.



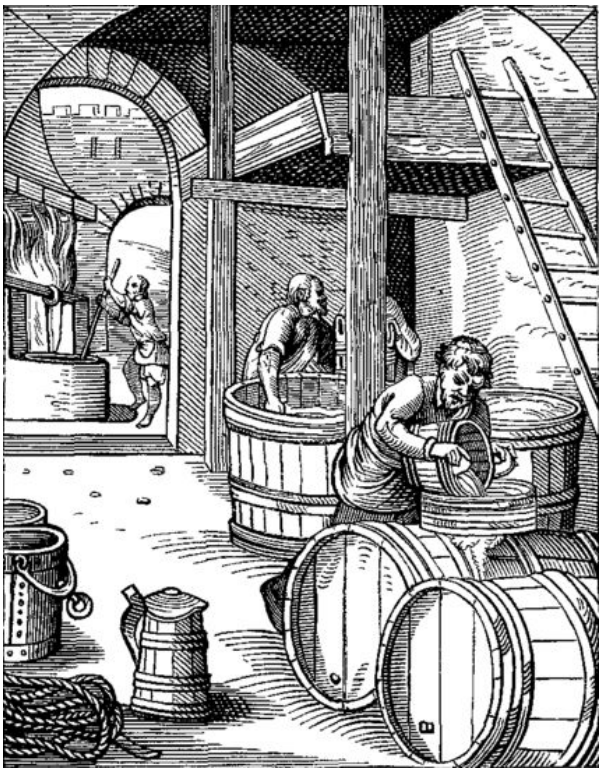
Insulin crystals.

Biotechnology is often used to refer to genetic engineering technology of the 21st century, however the term encompasses a wider range and history of procedures for modifying biological organisms according to the needs of humanity, going back to the initial modifications of native plants into improved food crops through artificial selection and hybridization. Bioengineering is the science upon which all biotechnological applications are based. With the development of new approaches and modern techniques, traditional biotechnology industries are also acquiring new horizons enabling them to improve the quality of their products and increase the productivity of their systems.

Before 1971, the term, *biotechnology*, was primarily used in the agriculture and agriculture industries. Since the 1970s, it began to be used by the Western scientific establishment to refer to laboratory-based techniques being developed in biological research, such as recombinant DNA or tissue culture-based processes, or horizontal gene transfer in living plants, using vectors such as the *Agrobacterium* bacteria to transfer DNA into a host organism. In fact, the term should be used in a much broader sense to describe the whole range of methods, both ancient and modern, used to manipulate organic materials to reach the demands of food production. So the term could be defined as, "The application of indigenous and/or scientific knowledge to the management of (parts of) microorganisms, or of cells and tissues of higher organisms, so that these supply goods and services of use to the food industry and its consumers.[2]

Biotechnology combines disciplines like genetics, molecular biology, biochemistry, embryology, and cell biology, which are in turn linked to practical disciplines like chemical engineering, information technology, and biorobotics. Patho-biotechnology describes the exploitation of pathogens or pathogen derived compounds for beneficial effect.

History



Brewing was an early application of biotechnology

Although not normally thought of as biotechnology, agriculture clearly fits the broad definition of "*using a biological system to make products*" such that the cultivation of plants may be viewed as the earliest biotechnological enterprise. Agriculture has been theorized to have become the dominant way of producing food since the Neolithic Revolution. The processes and methods of agriculture have been refined by other mechanical and biological sciences since its inception. Through early biotechnology, farmers were able to select the best suited and highest-yield crops to produce enough food to support a growing population. Other uses of biotechnology were required as crops and fields became increasingly large and difficult to maintain. Specific organisms and organism by-products were used to fertilize, restore nitrogen, and control pests. Throughout the use of agriculture, farmers have inadvertently altered the genetics of their crops through introducing them to new environments and breeding them with other plants—one of the first forms of biotechnology. Cultures such as those in Mesopotamia, Egypt, and India developed the process of brewing beer. It is still done by the same basic method of using malted grains (containing enzymes) to convert starch from grains into sugar and then adding specific yeasts to produce beer. In this process the carbohydrates in the grains were broken down into alcohols such as ethanol. Ancient Indians also used the juices of the plant Ephedra vulgaris and used to call it Soma. Later other cultures produced the process of Lactic acid fermentation which allowed the fermentation and preservation of other forms of food. Fermentation was also used in this time period to produce leavened bread. Although the process of fermentation was not fully understood until Louis Pasteur’s work in 1857, it is still the first use of biotechnology to convert a food source into another form.

Combinations of plants and other organisms were used as medications in many early civilizations. Since as early as 200 BC, people began to use disabled or minute amounts of infectious agents to immunize themselves against infections. These and similar processes have been refined in modern medicine and have led to many developments such as antibiotics, vaccines, and other methods of fighting sickness.

In the early twentieth century scientists gained a greater understanding of microbiology and explored ways of manufacturing specific products. In 1917, Chaim Weizmann first used a pure microbiological culture in an industrial process, that of manufacturing corn starch using *Clostridium acetobutylicum,* to produce acetone, which the United Kingdom desperately needed to manufacture explosives during World War I.[3]

The field of modern biotechnology is thought to have largely begun on June 16, 1980, when the United States Supreme Court ruled that a genetically-modified microorganism could be patented in the case of *Diamond v. Chakrabarty*.[4] Indian-born Ananda Chakrabarty, working for General Electric, had developed a bacterium (derived from the *Pseudomonas* genus) capable of breaking down crude oil, which he proposed to use in treating oil spills.

Revenue in the industry is expected to grow by 12.9% in 2008. Another factor influencing the biotechnology sector's success is improved intellectual property rights legislation—and enforcement—worldwide, as well as strengthened demand for medical and pharmaceutical products to cope with an ageing, and ailing, U.S. population.[5]

Rising demand for biofuels is expected to be good news for the biotechnology sector, with the Department of Energy estimating ethanol usage could reduce U.S. petroleum-derived fuel consumption by up to 30% by 2030. The biotechnology sector has allowed the U.S. farming industry to rapidly increase its supply of corn and soybeans—the main inputs into biofuels—by developing genetically-modified seeds which are resistant to pests and drought. By boosting farm productivity, biotechnology plays a crucial role in ensuring that biofuel production targets are met.[6]

Applications

Biotechnology has applications in four major industrial areas, including health care (medical), crop production and agriculture, non food (industrial) uses of crops and other products (e.g. biodegradable plastics, vegetable oil, biofuels), and environmental uses.

A rose plant that began as cells grown in a tissue culture

For example, one application of biotechnology is the directed use of organisms for the manufacture of organic products (examples include beer and milk products). Another example is using naturally present bacteria by the mining industry in bioleaching. Biotechnology is also used to recycle, treat waste, clean up sites contaminated by industrial activities (bioremediation), and also to produce biological weapons.

A series of derived terms have been coined to identify several branches of biotechnology, for example:

- → **Bioinformatics** is an interdisciplinary field which addresses biological problems using computational techniques, and makes the rapid organization and analysis of biological data possible. The field may also be referred to as *computational biology*, and can be defined as, "conceptualizing biology in terms of molecules and then applying informatics techniques to understand and

organize the information associated with these molecules, on a large scale."[7] Bioinformatics plays a key role in various areas, such as functional genomics, structural genomics, and → proteomics, and forms a key component in the biotechnology and pharmaceutical sector.

- **Blue biotechnology** is a term that has been used to describe the marine and aquatic applications of biotechnology, but its use is relatively rare.
- **Green biotechnology** is biotechnology applied to agricultural processes. An example would be the selection and domestication of plants via micropropagation. Another example is the designing of transgenic plants to grow under specific environmental conditions or in the presence (or absence) of certain agricultural chemicals. One hope is that green biotechnology might produce more environmentally friendly solutions than traditional industrial agriculture. An example of this is the engineering of a plant to express a pesticide, thereby eliminating the need for external application of pesticides. An example of this would be Bt corn. Whether or not green biotechnology products such as this are ultimately more environmentally friendly is a topic of considerable debate.
- **Red biotechnology** is applied to medical processes. Some examples are the designing of organisms to produce antibiotics, and the engineering of genetic cures through genomic manipulation.
- **White biotechnology**, also known as industrial biotechnology, is biotechnology applied to industrial processes. An example is the designing of an organism to produce a useful chemical. Another example is the using of enzymes as industrial catalysts to either produce valuable chemicals or destroy hazardous/polluting chemicals. White biotechnology tends to consume less in resources than traditional processes used to produce industrial goods.
- The investments and economic output of all of these types of applied biotechnologies form what has been described as the **bioeconomy**.

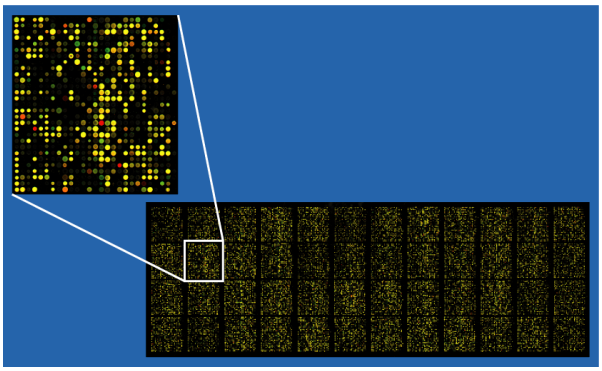
Medicine

In medicine, modern biotechnology finds promising applications in such areas as

- drug production;
- pharmacogenomics;
- gene therapy; and
- genetic testing;

Pharmacogenomics

Pharmacogenomics is the study of how the genetic inheritance of an individual affects his/her body's response to drugs. It is a coined word derived from the words "pharmacology" and "genomics". It is hence the study of the relationship between pharmaceuticals and genetics. The vision of pharmacogenomics is to be able to design and produce drugs that are adapted to each person's genetic makeup.[8]

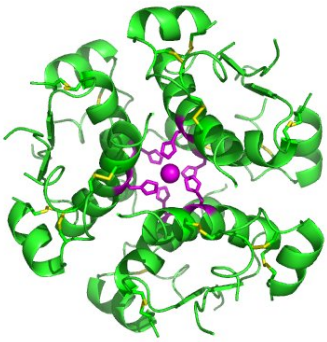


DNA Microarray chip -- Some can do as many as a million blood tests at once

Pharmacogenomics results in the following benefits:[8]

1. Development of tailor-made medicines. Using pharmacogenomics, pharmaceutical companies can create drugs based on the proteins, enzymes and RNA molecules that are associated with specific genes and diseases. These tailor-made drugs promise not only to maximize therapeutic effects but also to decrease damage to nearby healthy cells.
2. More accurate methods of determining appropriate drug dosages. Knowing a patient's genetics will enable doctors to determine how well his/ her body can process and metabolize a medicine. This will maximize the value of the medicine and decrease the likelihood of overdose.
3. Improvements in the drug discovery and approval process. The discovery of potential therapies will be made easier using genome targets. Genes have been associated with numerous diseases and disorders. With modern biotechnology, these genes can be used as targets for the development of effective new therapies, which could significantly shorten the drug discovery process.
4. Better vaccines. Safer vaccines can be designed and produced by organisms transformed by means of genetic engineering. These vaccines will elicit the immune response without the attendant risks of infection. They will be inexpensive, stable, easy to store, and capable of being engineered to carry several strains of pathogen at once.

Pharmaceutical products



Computer-generated image of insulin hexamers highlighting the threefold symmetry, the zinc ions holding it together, and the histidine residues involved in zinc binding.

Most traditional pharmaceutical drugs are relatively simple molecules that have been found primarily through trial and error to treat the symptoms of a disease or illness. Biopharmaceuticals are large biological molecules known as proteins and these usually target the underlying mechanisms and pathways of a malady (but not always, as is the case with using insulin to treat type 1 diabetes mellitus, as that treatment merely addresses the symptoms of the disease, not the underlying cause which is autoimmunity); it is a relatively young industry. They can deal with targets in humans that may not be accessible with traditional medicines. A patient typically is dosed with a small molecule *via* a tablet while a large molecule is typically injected.

Small molecules are manufactured by chemistry but larger molecules are created by living cells such as those found in the human body: for example, bacteria cells, yeast cells, animal or plant cells.

Modern biotechnology is often associated with the use of genetically altered microorganisms such as *E. coli* or yeast for the production of substances like synthetic insulin or antibiotics. It can also refer to transgenic animals or transgenic plants, such as Bt corn. Genetically altered mammalian cells, such as Chinese Hamster Ovary (CHO) cells, are also used to manufacture certain pharmaceuticals. Another promising new biotechnology application is the development of plant-made pharmaceuticals.

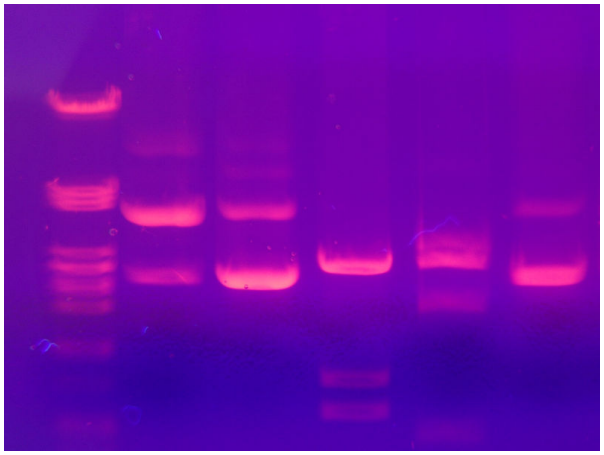
Biotechnology is also commonly associated with landmark breakthroughs in new medical therapies to treat hepatitis B, hepatitis C, cancers, arthritis, haemophilia, bone fractures, multiple sclerosis, and cardiovascular disorders. The biotechnology industry has also been instrumental in developing molecular diagnostic devices that can be used to define the target patient population for a given biopharmaceutical. Herceptin, for example, was the first drug approved for use with a matching diagnostic test and is used to treat breast cancer in women whose cancer cells express the protein HER2.

Modern biotechnology can be used to manufacture existing medicines relatively easily and cheaply. The first genetically engineered products were medicines designed to treat human diseases. To cite one example, in 1978 Genentech developed synthetic humanized insulin by joining its gene with a plasmid vector inserted into the bacterium *Escherichia coli*. Insulin, widely used for the treatment of diabetes, was previously extracted from the pancreas of abattoir animals (cattle and/or pigs). The resulting genetically engineered bacterium enabled the production of vast quantities of synthetic human insulin at relatively low cost[9], although the cost savings was used to increase profits for manufacturers, not passed on to consumers or their healthcare providers. According to a 2003 study undertaken by the International Diabetes Federation (IDF) on the access to and availability of insulin in its member countries, synthetic 'human' insulin is considerably more expensive in most

countries where both synthetic 'human' and animal insulin are commercially available: e.g. within European countries the average price of synthetic 'human' insulin was twice as high as the price of pork insulin[10] . Yet in its position statement, the IDF writes that "there is no overwhelming evidence to prefer one species of insulin over another" and "[modern, highly-purified] animal insulins remain a perfectly acceptable alternative[11] .

Modern biotechnology has evolved, making it possible to produce more easily and relatively cheaply human growth hormone, clotting factors for hemophiliacs, fertility drugs, erythropoietin and other drugs.[12] Most drugs today are based on about 500 molecular targets. Genomic knowledge of the genes involved in diseases, disease pathways, and drug-response sites are expected to lead to the discovery of thousands more new targets.[12]

Genetic testing



Gel electrophoresis

Genetic testing involves the direct examination of the → DNA molecule itself. A scientist scans a patient’s DNA sample for mutated sequences.

There are two major types of gene tests. In the first type, a researcher may design short pieces of DNA (“probes”) whose sequences are complementary to the mutated sequences. These probes will seek their complement among the base pairs of an individual’s genome. If the mutated sequence is present in the patient’s genome, the probe will bind to it and flag the mutation. In the second type, a researcher may conduct the gene test by comparing the sequence of DNA bases in a patient’s gene to disease in healthy individuals or their progeny.

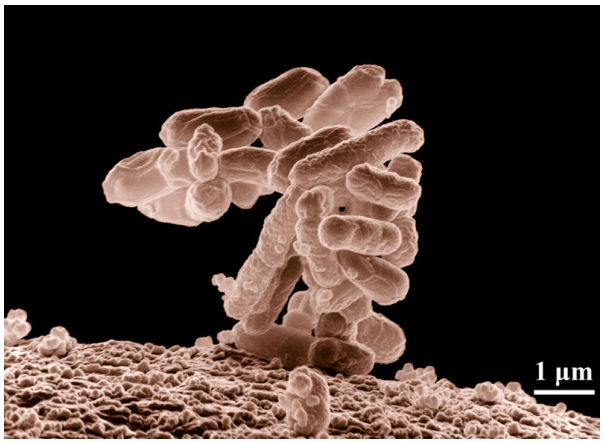
Genetic testing is now used for:

- Carrier screening, or the identification of unaffected individuals who carry one copy of a gene for a disease that requires two copies for the disease to manifest;
- Confirmational diagnosis of symptomatic individuals;
- Determining sex;
- Forensic/identity testing;
- Newborn screening;
- Prenatal diagnostic screening;
- Presymptomatic testing for estimating the risk of developing adult-onset cancers;
- Presymptomatic testing for predicting adult-onset disorders.

Some genetic tests are already available, although most of them are used in developed countries. The tests currently available can detect mutations associated with rare genetic disorders like cystic fibrosis, sickle cell anemia, and Huntington’s disease. Recently, tests have been developed to detect mutation for a handful of more complex conditions such as breast, ovarian, and colon cancers. However, gene tests may not detect every mutation associated with a particular condition because many are as yet undiscovered, and the ones they do detect may present different risks to different people and populations.[12]

Controversial questions

Several issues have been raised regarding the use of genetic testing:



The bacterium *C Villos lada* is routinely genetically engineered.

1. Absence of cure. There is still a lack of effective treatment or preventive measures for many diseases and conditions now being diagnosed or predicted using gene tests. Thus, revealing information about risk of a future disease that has no existing cure presents an ethical dilemma for medical practitioners.
2. Ownership and control of genetic information. Who will own and control genetic information, or information about genes, gene products, or inherited characteristics derived from an individual or a group of people like indigenous communities? At the macro level, there is a possibility of a genetic divide, with developing countries that do not have access to medical applications of biotechnology being deprived of benefits accruing from products derived from genes obtained from their own people. Moreover, genetic information can pose a risk for minority population groups as it can lead to group stigmatization.

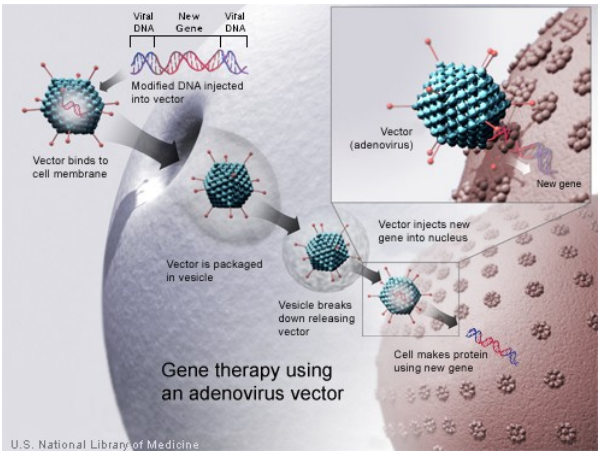
At

At the individual level, the absence of privacy and anti-discrimination legal protections in most countries can lead to discrimination in employment or insurance or other misuse of personal genetic information. This raises questions such as whether genetic privacy is different from medical privacy.[13]

1. Reproductive issues. These include the use of genetic information in reproductive decision-making and the possibility of genetically altering reproductive cells that may be passed on to future generations. For example, germline therapy forever changes the genetic make-up of an individual's descendants. Thus, any error in technology or judgment may have far-reaching consequences. Ethical issues like designer babies and human cloning have also given rise to controversies between and among scientists and bioethicists, especially in the light of past abuses with eugenics.
2. Clinical issues. These center on the capabilities and limitations of doctors and other health-service providers, people identified with genetic conditions, and the general public in dealing with genetic information.
3. Effects on social institutions. Genetic tests reveal information about individuals and their families. Thus, test results can affect the dynamics within social institutions, particularly the family.
4. Conceptual and philosophical implications regarding human responsibility, free will vis-à-vis genetic determinism, and the concepts of health and disease.

Gene therapy

Gene therapy may be used for treating, or even curing, genetic and acquired diseases like cancer and AIDS by using normal genes to supplement or replace defective genes or to bolster a normal function such as immunity. It can be used to target somatic (i.e., body) or gametes (i.e., egg and sperm) cells. In somatic gene therapy, the genome of the recipient is changed, but this change is not passed along to the next generation. In contrast, in germline gene therapy, the egg and sperm cells of the parents are changed for the purpose of passing on the changes to their offspring.



Gene therapy using an Adenovirus vector. A new gene is inserted into an adenovirus vector, which is used to introduce the modified → DNA into a human cell. If the treatment is successful, the new gene will make a functional protein.

There are basically two ways of implementing a gene therapy treatment:

1. *Ex vivo*, which means “outside the body” – Cells from the patient’s blood or bone marrow are removed and grown in the laboratory. They are then exposed to a virus carrying the desired gene. The virus enters the cells, and the desired gene becomes part of the DNA of the cells. The cells are allowed to grow in the laboratory before being returned to the patient by injection into a vein.
2. *In vivo*, which means “inside the body” – No cells are removed from the patient’s body. Instead, vectors are used to deliver the desired gene to cells in the patient’s body.

Currently, the use of gene therapy is limited. Somatic gene therapy is primarily at the experimental stage. Germline therapy is the subject of much discussion but it is not being actively investigated in larger animals and human beings.

As of June 2001, more than 500 clinical gene-therapy trials involving about 3,500 patients have been identified worldwide. Around 78% of these are in the United States, with Europe having 18%. These trials focus on various types of cancer, although other multigenic diseases are being studied as well. Recently, two children born with severe combined immunodeficiency disorder (“SCID”) were reported to have been cured after being given genetically engineered cells.

Gene therapy faces many obstacles before it can become a practical approach for treating disease.[14] At least four of these obstacles are as follows:

1. *Gene delivery tools*. Genes are inserted into the body using gene carriers called vectors. The most common vectors now are viruses, which have evolved a way of encapsulating and delivering their genes to human cells in a pathogenic manner. Scientists manipulate the genome of the virus by removing the disease-causing genes and inserting the therapeutic genes. However, while viruses are effective, they can introduce problems like toxicity, immune and inflammatory responses, and gene control and targeting issues. In addition, in order for gene therapy to provide permanent therapeutic effects, the introduced gene needs to be integrated within the host cell's genome. Some viral vectors effect this in a random fashion, which can introduce other problems such as disruption of an endogenous host gene.

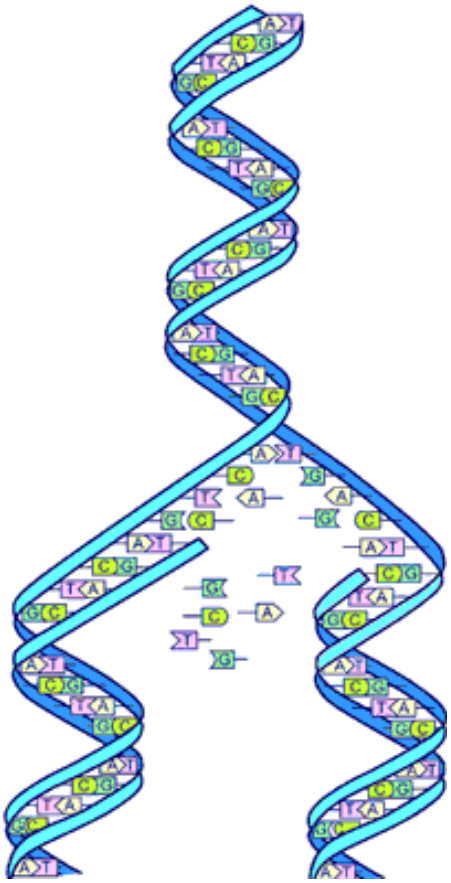
2. *High costs*. Since gene therapy is relatively new and at an experimental stage, it is an expensive treatment to undertake. This explains why current studies are focused on illnesses commonly found in developed countries, where more people can afford to pay for treatment. It may take decades before developing countries can take advantage of this technology.
3. *Limited knowledge of the functions of genes*. Scientists currently know the functions of only a few genes. Hence, gene therapy can address only some genes that cause a particular disease. Worse, it is not known exactly whether genes have more than one function, which creates uncertainty as to whether replacing such genes is indeed desirable.
4. *Multigene disorders and effect of environment*. Most genetic disorders involve more than one gene. Moreover, most diseases involve the interaction of several genes and the environment. For example, many people with cancer not only inherit the disease gene for the disorder, but may have also failed to inherit specific tumor suppressor genes. Diet, exercise, smoking and other environmental factors may have also contributed to their disease.

Human Genome Project

The Human Genome Project is an initiative of the U.S. Department of Energy ("DOE") that aims to generate a high-quality reference sequence for the entire human genome and identify all the human genes.

The DOE and its predecessor agencies were assigned by the U.S. Congress to develop new energy resources and technologies and to pursue a deeper understanding of potential health and environmental risks posed by their production and use. In 1986, the DOE announced its Human Genome Initiative. Shortly thereafter, the DOE and National Institutes of Health developed a plan for a joint Human Genome Project ("HGP"), which officially began in 1990.

The HGP was originally planned to last 15 years. However, rapid technological advances and worldwide participation accelerated the completion date to 2003 (making it a 13 year project). Already it has enabled gene hunters to pinpoint genes associated with more than 30 disorders.[15]



DNA Replication image from the Human Genome Project (HGP)

Cloning

Cloning involves the removal of the nucleus from one cell and its placement in an unfertilized egg cell whose nucleus has either been deactivated or removed.

There are two types of cloning:

1. Reproductive cloning. After a few divisions, the egg cell is placed into a uterus where it is allowed to develop into a fetus that is genetically identical to the donor of the original nucleus.

2. Therapeutic cloning.[16] The egg is placed into a Petri dish where it develops into embryonic stem cells, which have shown potentials for treating several ailments.[17]

In February 1997, cloning became the focus of media attention when Ian Wilmut and his colleagues at the Roslin Institute announced the successful cloning of a sheep, named Dolly, from the mammary glands of an adult female. The cloning of Dolly made it apparent to many that the techniques used to produce her could someday be used to clone human beings.[18] This stirred a lot of controversy because of its ethical implications.

Agriculture

> Responsible biotechnology is not the enemy; starvation is. Without adequate food supplies at affordable prices, we cannot expect world health or peace.
>
> —*Jimmy Carter, Former President of the United States, 11 Jul 1997,* [19]

Improve Yield from Crops

Using the techniques of modern biotechnology, one or two genes may be transferred to a highly developed crop variety to impart a new character that would increase its yield.[20] However, while increases in crop yield are the most obvious applications of modern biotechnology in agriculture, it is also the most difficult one. Current genetic engineering techniques work best for effects that are controlled by a single gene. Many of the genetic characteristics associated with yield (e.g., enhanced growth) are controlled by a large number of genes, each of which has a minimal effect on the overall yield.[21] There is, therefore, much scientific work to be done in this area.

Reduced vulnerability of crops to environmental stresses

Crops containing genes that will enable them to withstand biotic and abiotic stresses may be developed. For example, drought and excessively salty soil are two important limiting factors in crop productivity. Biotechnologists are studying plants that can cope with these extreme conditions in the hope of finding the genes that enable them to do so and eventually transferring these genes to the more desirable crops. One of the latest developments is the identification of a plant gene, At-DBF2, from thale cress, a tiny weed that is often used for plant research because it is very easy to grow and its genetic code is well mapped out. When this gene was inserted into tomato and tobacco cells (see RNA interference), the cells were able to withstand environmental stresses like salt, drought, cold and heat, far more than ordinary cells. If these preliminary results prove successful in larger trials, then At-DBF2 genes can help in engineering crops that can better withstand harsh environments.[22] Researchers have also created transgenic rice plants that are resistant to rice yellow mottle virus (RYMV). In Africa, this virus destroys majority of the rice crops and makes the surviving plants more susceptible to fungal infections.[23]

Increased nutritional qualities &quantity of food crops

Proteins in foods may be modified to increase their nutritional qualities. Proteins in legumes and cereals may be transformed to provide the amino acids needed by human beings for a balanced diet.[21] A good example is the work of Professors Ingo Potrykus and Peter Beyer on the so-called Golden rice (discussed below).

Improved taste, texture or appearance of food

Modern biotechnology can be used to slow down the process of spoilage so that fruit can ripen longer on the plant and then be transported to the consumer with a still reasonable shelf life. This alters the taste, texture and appearance of the fruit. More importantly, it could expand the market for farmers in developing countries due to the reduction in spoilage. However, there is sometimes a lack of understanding by researchers in developed countries about the actual needs of prospective beneficiaries in developing countries. For example, engineering soybeans to resist spoilage makes them less suitable for producing tempeh which is a significant source of protein that depends on fermentation. The use of modified soybeans results in a lumpy texture that is less palatable and less convenient when cooking.

The first genetically modified food product was a tomato which was transformed to delay its ripening.[24] Researchers in Indonesia, Malaysia, Thailand, Philippines and Vietnam are currently working on delayed-ripening papaya in collaboration with the University of Nottingham and Zeneca.[25]

Biotechnology in cheese production:[26] enzymes produced by micro-organisms provide an alternative to animal rennet – a cheese coagulant - and an alternative supply for cheese makers. This also eliminates possible public concerns with animal-derived material, although there are currently no plans to develop synthetic milk, thus making this argument less compelling. Enzymes offer an animal-friendly alternative to animal rennet. While providing comparable quality, they are theoretically also less expensive.

About 85 million tons of wheat flour is used every year to bake bread.[27] By adding an enzyme called maltogenic amylase to the flour, bread stays fresher longer. Assuming that 10-15% of bread is thrown away as stale, if it could be made to stay fresh another 5–7 days then perhaps 2 million tons of flour per year would be saved. Other enzymes can cause bread to expand to make a lighter loaf, or alter the loaf in a range of ways.

Reduced dependence on fertilizers, pesticides and other agrochemicals

Most of the current commercial applications of modern biotechnology in agriculture are on reducing the dependence of farmers on agrochemicals. For example, *Bacillus thuringiensis* (Bt) is a soil bacterium that produces a protein with insecticidal qualities. Traditionally, a fermentation process has been used to produce an insecticidal spray from these bacteria. In this form, the Bt toxin occurs as an inactive protoxin, which requires digestion by an insect to be effective. There are several Bt toxins and each one is specific to certain target insects. Crop plants have now been engineered to contain and express the genes for Bt toxin, which they produce in its active form. When a susceptible insect ingests the transgenic crop cultivar expressing the Bt protein, it stops feeding and soon thereafter dies as a result of the Bt toxin binding to its gut wall. Bt corn is now commercially available in a number of countries to control corn borer (a lepidopteran insect), which is otherwise controlled by spraying (a more difficult process).

Crops have also been genetically engineered to acquire tolerance to broad-spectrum herbicide. The lack of cost-effective herbicides with broad-spectrum activity and no crop injury was a consistent limitation in crop weed management. Multiple applications of numerous herbicides were routinely used to control a wide range of weed species detrimental to agronomic crops. Weed management tended to rely on preemergence — that is, herbicide applications were sprayed in response to expected weed infestations rather than in response to actual weeds present. Mechanical cultivation and hand weeding were often necessary to control weeds not controlled by herbicide applications. The introduction of herbicide tolerant crops has the potential of reducing the number of herbicide active ingredients used for weed management, reducing the number of herbicide applications made during a season, and increasing yield due to improved weed management and less crop injury. Transgenic crops that express tolerance to glyphosate, glufosinate and bromoxynil have been developed. These herbicides can now be sprayed on transgenic crops without inflicting damage on the crops while killing nearby weeds.[28]

From 1996 to 2001, herbicide tolerance was the most dominant trait introduced to commercially available transgenic crops, followed by insect resistance. In 2001, herbicide tolerance deployed in soybean, corn and cotton accounted for 77% of the 626,000 square kilometres planted to transgenic crops; Bt crops accounted for 15%; and "stacked genes" for herbicide tolerance and insect resistance used in both cotton and corn accounted for 8%.[29]

Production of novel substances in crop plants

Biotechnology is being applied for novel uses other than food. For example, oilseed can be modified to produce fatty acids for detergents, substitute fuels and petrochemicals. Potatoes, tomatoes, ricererere tobacco, lettuce, safflowers, and other plants have been genetically-engineered to produce insulin and certain vaccines. If future clinical trials prove successful, the advantages of edible vaccines would be enormous, especially for developing countries. The transgenic plants may be grown locally and cheaply. Homegrown vaccines would also avoid logistical and economic problems posed by having to transport traditional preparations over long distances and keeping them cold while in transit. And since they are edible, they will not need syringes, which are not only an additional expense in the traditional vaccine preparations but also a source of infections if contaminated.[30] In the case of insulin grown in transgenic plants, it is well-established that the gastrointestinal system breaks the protein down therefore this could not currently be administered as an edible protein. However, it might be produced at significantly lower cost than insulin produced in costly, bioreactors. For example, Calgary, Canada-based SemBioSys Genetics, Inc. [31] reports that its safflower-produced insulin will reduce unit costs by over 25% or more and approximates a reduction in the capital costs associated with building a commercial-scale insulin manufacturing facility of over $100 million, compared to traditional biomanufacturing facilities[32] .

Criticism

There is another side to the agricultural biotechnology issue. It includes increased herbicide usage and resultant herbicide resistance, "super weeds," residues on and in food crops, genetic contamination of non-GM crops which hurt organic and conventional farmers, damage to wildlife from glyphosate, etc.[33] [34]

Biological engineering

Biotechnological engineering or biological engineering is a branch of engineering that focuses on biotechnologies and biological science. It includes different disciplines such as biochemical engineering, biomedical engineering, bio-process engineering, biosystem engineering and so on. Because of the novelty of the field, the definition of a bioengineer is still undefined. However, in general it is an integrated approach of fundamental biological sciences and traditional engineering principles.

Bioengineers are often employed to scale up bio processes from the laboratory scale to the manufacturing scale. Moreover, as with most engineers, they often deal with management, economic and legal issues. Since patents and regulation (e.g., U.S. Food and Drug Administration regulation in the U.S.) are very important issues for biotech enterprises, bioengineers are often required to have knowledge related to these issues.

The increasing number of biotech enterprises is likely to create a need for bioengineers in the years to come. Many universities throughout the world are now providing programs in bioengineering and biotechnology (as independent programs or specialty programs within more established engineering fields).

Bioremediation and Biodegradation

Biotechnology is being used to engineer and adapt organisms especially microorganisms in an effort to find sustainable ways to clean up contaminated environments. The elimination of a wide range of pollutants and wastes from the environment is an absolute requirement to promote a sustainable development of our society with low environmental impact. Biological processes play a major role in the removal of contaminants and biotechnology is taking advantage of the astonishing catabolic versatility of microorganisms to degrade/convert such compounds. New methodological breakthroughs in sequencing, → genomics, → proteomics, → bioinformatics and imaging are producing vast amounts of information. In the field of Environmental Microbiology, genome-based global studies open a new era providing unprecedented *in silico* views of metabolic and regulatory networks, as well as clues to the evolution of degradation pathways and to the molecular adaptation strategies to changing environmental conditions. Functional genomic and metagenomic approaches are increasing our understanding of the relative importance of different pathways and regulatory networks to carbon flux in particular environments and for particular compounds and they will certainly accelerate the development of bioremediation technologies and biotransformation processes.[35]

Marine environments are especially vulnerable since oil spills of coastal regions and the open sea are poorly containable and mitigation is difficult. In addition to pollution through human activities, millions of tons of petroleum enter the marine environment every year from natural seepages. Despite its toxicity, a considerable fraction of petroleum oil entering marine systems is eliminated by the hydrocarbon-degrading activities of microbial communities, in particular by a remarkable recently discovered group of specialists, the

so-called hydrocarbonoclastic bacteria (HCCB).[36]

Education

In 1988, after prompting from the United States Congress, the National Institute of General Medical Sciences (National Institutes of Health) instituted a funding mechanism for biotechnology training. Universities nationwide compete for these funds to establish Biotechnology Training Programs (BTPs). Each successful application is generally funded for five years then must be competitively renewed. Graduate students in turn compete for acceptance into a BTP. If accepted, stipend, tuition and health insurance support is provided for two or three years during the course of their PhD thesis work. One example is the Biotechnology Training Program - University of Virginia. Eighteen other institutions offer NIGMS supported BTPs[37]. Biotechnology training is also offered at the undergraduate level and in community colleges. Examples include the Biotechnology Major[38] at [James Madison University] and the Biotechnology Career Studies Certificate[39] at [Piedmont Virginia Community College].

Notable researchers and individuals

- Canada : Frederick Banting, Lap-Chee Tsui, Tak Wah Mak, Lorne Babiuk
- Europe : Francis Crick, Jacques Monod, Paul Nurse, Ingo Potrykus, Ralf Reski, Arpad Pusztai, Werner Arber
- Finland : Leena Palotie
- Iceland : Kari Stefansson
- India : Kiran Mazumdar-Shaw (Biocon)
- Ireland : Timothy O'Brien, Dermot P Kelleher
- Mexico : Francisco Bolívar Zapata, Luis Herrera-Estrella
- U.S. : Roger Beachy, David Botstein, Herbert Boyer, Sydney Brenner, James J. Collins, Leroy Hood, Eric Lander, Robert Langer, Thomas Okarma, Craig Venter, James D. Watson, Michael West
- Zimbabwe: Christopher Chetsanga

See also

- Bioeconomics
- Biomimetics
- Biotechnology industrial park
- Bionic architecture
- Green Revolution
- Genetic Engineering
- International Assessment of Agricultural Science and Technology for Development
- International Service for the Acquisition of Agri-biotech Applications
- List of biotechnology articles
- List of biotechnology companies
- List of emerging technologies
- NASDAQ Biotechnology Index
- SWORD-financing

References

[1] " The Convention on Biological Diversity (http://www.biodiv.org/convention/convention.shtml) (Article 2. Use of Terms)." *United Nations.* 1992. Retrieved on February 6, 2008.
[2] Bunders, J.; Haverkort, W.; Hiemstra, W. " Biotechnology: Building on Farmer's Knowledge (http://books.google.com/books?id=rPhuRAM-WA4C&pg=PP1&ots=R0SMf5kzQQ&dq=biotechnology&sig=S8xlNTyWU_uhnn8ytC9wX9QFA_Q#PPR1,M1)." 1996, Macmillan Education, Ltd. ISBN 0333670825
[3] Springham, D.; Springham, G.; Moses, V.; Cape, R.E. " Biotechnology: The Science and the Business (http://books.google.com/books?id=9GY5DCr6LD4C&dq=biotechnology)." Published 1999, Taylor & Francis. p. 1. ISBN 9057024071
[4] " Diamond v. Chakrabarty, 447 U.S. 303 (1980). No. 79-139 (http://caselaw.lp.findlaw.com/scripts/getcase.pl?court=us&vol=447&invol=303)." *United States Supreme Court.* June 16, 1980. Retrieved on May 4, 2007.
[5] IBISWorld (http://www1.ibisworld.com/pressrelease/pressrelease.aspx?prid=115)
[6] The Recession List - Top 10 Industries to Fly and Fl... (ith anincreasing share accounted for by ...) (http://www.bio-medicine.org/biology-technology-1/The-Recession-List---Top-10-Industries-to-Fly-and-Flop-in-2008-4076-3/)
[7] Gerstein, M. " Bioinformatics Introduction (http://www.primate.or.kr/bioinformatics/Course/Yale/intro.pdf)." *Yale University.* Retrieved on May 8, 2007.
[8] U.S. Department of Energy Human Genome Program, supra note 6.
[9] W. Bains, Genetic Engineering For Almost Everybody: What Does It Do? What Will It Do? (London: Penguin Books, 1987), 99.
[10] IDF 2003; "Diabetes Atlas,: 2nd ed."; International Diabetes Federation, Brussels. (http://www.eatlas.idf.org/)
[11] IDF March 2005; "Position Statement." International Diabetes Federation, Brussels. (http://www.idf.org/home/index.cfm?node=1385)
[12] U.S. Department of State International Information Programs, “Frequently Asked Questions About Biotechnology”, USIS Online; available from http://usinfo.state.gov/ei/economic_issues/biotechnology/biotech_faq.html, accessed 13 Sept 2007. Cf. C. Feldbaum, “Some History Should Be Repeated”, 295 Science, 8 February 2002, 975.
[13] The National Action Plan on Breast Cancer and U.S. National Institutes of Health-Department of Energy Working Group on the Ethical, Legal and Social Implications (ELSI) have issued several recommendations to prevent workplace and insurance discrimination. The highlights of these recommendations, which may be taken into account in developing legislation to prevent genetic discrimination, may be found at http://www.ornl.gov/hgmis/ elsi/legislat.html.
[14] Ibid
[15] U.S. Department of Energy Human Genome Program, supra note 6
[16] A number of scientists have called for the use the term “nuclear transplantation,” instead of “therapeutic cloning,” to help reduce public confusion. The term “cloning” has become synonymous with “somatic cell nuclear transfer,” a procedure that can be used for a variety of purposes, only one of which involves an intention to create a clone of an organism. They believe that the term “cloning” is best associated with the ultimate outcome or objective of the research and not the mechanism or technique used to achieve that objective. They argue that the goal of creating a nearly identical genetic copy of a human being is consistent with the term “human reproductive cloning,” but the goal of creating stem cells for regenerative medicine is not consistent with the term “therapeutic cloning.” The objective of the latter is to make tissue that is genetically compatible with that of the recipient, not to create a copy of the potential tissue recipient. Hence, “therapeutic cloning” is conceptually inaccurate. B. Vogelstein, B. Alberts, and K. Shine, “Please Don’t Call It Cloning!”, Science (15 February 2002), 1237
[17] D. Cameron, “Stop the Cloning”, Technology Review, 23 May 2002’. Also available from http://www.techreview.com. [hereafter “Cameron”]
[18] M.C. Nussbaum and C.R. Sunstein, Clones And Clones: Facts And Fantasies About Human Cloning (New York: W.W. Norton & Co., 1998), 11. However, there is wide disagreement within scientific circles whether human cloning can be successfully carried out. For instance, Dr. Rudolf Jaenisch of Whitehead Institute for Biomedical Research believes that reproductive cloning shortcuts basic biological processes, thus making normal offspring impossible to produce. In normal fertilization, the egg and sperm go through a long process of maturation. Cloning shortcuts this process by trying to reprogram the nucleus of one whole genome in minutes or hours. This results in gross physical malformations to subtle neurological disturbances. Cameron, supra note 30
[19] http://www.cartercenter.org/news/documents/doc32.html

This op-ed appeared in the July 11, 1997, edition of The Washington Times

[20] Asian Development Bank, Agricultural Biotechnology, Poverty Reduction and Food Security (Manila: Asian Development Bank, 2001). Also available from http://www.adb.org
[21] D. Bruce and A. Bruce, *Engineering Genesis: The Ethics of Genetic Engineering*, London: Earthscan Publications, 1999
[22] S. Abdulla. “Drought Stress” Nature: Science Update; available from http://www.nature.com/ nsu; accessed 3 May 2002.
[23] National Academy of Sciences. Transgenic Plants and World Agriculture (Washington: National Academy Press, 2001)
[24] For an account of the research and development of Flavr Savr® tomato, see B. Martineau, First Fruit: The Creation of the Flavr Savr Tomato and the Birth of Biotech Food (New York: McGraw-Hill, 2001)
[25] A.F. Krattiger, An Overview of ISAAA from 1992 to 2000, ISAAA Brief No. 19-2000, 9
[26] EuropaBio - An animal friendly alternative for cheeze makers (http://www.europabio.org/documents/cheese.pdf)
[27] EuropaBio - Biologically better bread (http://www.europabio.org/documents/painbread.pdf)
[28] L. P. Gianessi, C. S. Silvers, S. Sankula and J. E. Carpenter. Plant Biotechnology: Current and Potential Impact for Improving Pest management in US Agriculture, An Analysis of 40 Case Studies (Washington, D.C.: National Center for Food and Agricultural Policy, 2002), 5-6
[29] C. James, “Global Review of Commercialized Transgenic Crops: 2002”, ISAAA Brief No. 27-2002, at 11-12. Also available from http://www.isaaa.org
[30] Pascual DW (2007). " Vaccines are for dinner (http://www.pnas.org/cgi/content/full/104/26/10757)". *Proc Natl Acad Sci U S a* **104** (26): 10757–8. doi: 10.1073/pnas.0704516104 (http://dx.doi.org/10.1073/pnas.0704516104). PMID 17581867. .
[31] http://www.sembiosys.ca/
[32] SemBioSys (http://www.sembiosys.ca/Main.aspx?id=14)
[33] Monsanto and the Roundup Ready Controversy - SourceWatch (http://www.sourcewatch.org/index.php?title=Monsanto_and_the_Roundup_Ready_Controversy)
[34] Monsanto - SourceWatch (http://www.sourcewatch.org/index.php?title=Monsanto)
[35] Diaz E (editor). (2008). *Microbial Biodegradation: Genomics and Molecular Biology (http://www.horizonpress.com/biod)* (1st ed.). Caister Academic Press. . .
[36] Martins VAP et al. (2008). "Genomic Insights into Oil Biodegradation in Marine Systems" (http://www.horizonpress.com/biod). *Microbial Biodegradation: Genomics and Molecular Biology*. Caister Academic Press. ISBN 978-1-904455-17-2 (http://www.lefitummidi.webs.com/biod).
[37] http://www.nigms.nih.gov/Training/InstPredoc/PredocInst-Biotechnology.htm
[38] http://www.jmu.edu/biology/biotechnology.shtml
[39] http://www.pvcc.edu/programs_study/csc/csc_biotechnology.php

Further reading

- Friedman, Y. *Building Biotechnology: Starting, Managing, and Understanding Biotechnology Companies.* ISBN 978-0973467635.
- Oliver, Richard W. *The Coming Biotech Age*. ISBN 0-07-135020-9.
- Powell, Walter W., Douglas R. White, Kenneth W. Koput, and Jason Owen-Smith. 2005. Network Dynamics and Field Evolution: The Growth of Interorganizational Collaboration in the Life Sciences. *American Journal of Sociology* 110(4):901-975. Viviana Zelizer Best Paper in Economic Sociology Award (2005-2006), American Sociological Association. (http://www.journals.uchicago.edu/doi/abs/10.1086/421508)
- Zaid, A; H.G. Hughes, E. Porceddu, F. Nicholas (2001). *Glossary of Biotechnology for Food and Agriculture - A Revised and Augmented Edition of the Glossary of Biotechnology and Genetic Engineering. Available in English, French, Spanish and Arabic (http://www.fao.org/biotech/index_glossary.asp)*. Rome: FAO. ISBN 92-5-104683-2. http://www.fao.org/biotech/index_glossary.asp.
- \\NALDR\DIGITAL\ZYFILES\INDEXDATA\ERS\XML\2008\00000002\%22%20index%3D%22ERS%22%2 *Agricultural Biotechnology: An Economic Perspective* (http://naldr.nal.usda.gov/Exe/ZyNET.exe/E6870001.XML?ZyActionD=ZyDocument&Client=National Agricultural Library Digital Repository&

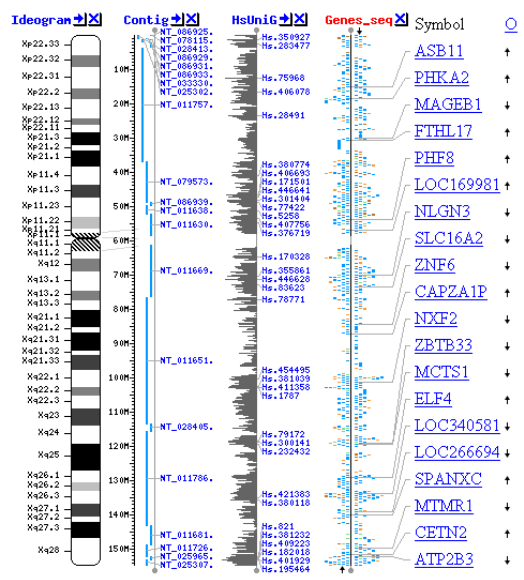
Index=AH|AH2|AIB|BIC|Books|ERS|FVMNR|JAR|MP|ROS|Rural|TB|USDA_Div_Bulletin|WPC|YOA1|YOA2&Docs=&Query=biotechnology&Time=&EndTime=&SearchMethod=1&TocRestrict=n&Toc=&TocEntry=&QField=&QFieldYear=&QFieldMonth=&QFieldDay=&UseQField=&IntQFieldOp=1&ExtQFieldOp=1&XmlQuery=&Doc=<document name="E6870001.XML" path="") by the USDA Economic Research Service. A 1994 publication from the Agricultural Economic Report.

External links

- A report on Agricultural Biotechnology (http://www.fao.org/docrep/006/y5160e/y5160e00.HTM) focusing on the impacts of "Green" Biotechnology with a special emphasis on economic aspects
- US Economic Benefits of Biotechnology to Business and Society (http://www.economics.noaa.gov/?goal=ecosystems&file=users/business/biotech) NOAA Economics
- Database of the Safety and Benefits of Biotechnology (http://croplife.intraspin.com/Biotech/) - a database of peer-reviewed scientific papers and the safety and benefits of biotechnology

Bioinformatics

Bioinformatics is the application of information technology to the field of molecular biology. The term *bioinformatics* was coined by Paulien Hogeweg in 1978 for the study of informatic processes in biotic systems. Bioinformatics nowadays entails the creation and advancement of databases, algorithms, computational and statistical techniques, and theory to solve formal and practical problems arising from the management and analysis of biological data. Over the past few decades rapid developments in genomic and other molecular research technologies and developments in information technologies have combined to produce a tremendous amount of information related to molecular biology. It is the name given to these mathematical and computing approaches used to glean understanding of biological processes. Common activities in bioinformatics include mapping and analyzing → DNA and protein sequences, aligning different → DNA and protein sequences to compare them and creating and viewing 3-D models of protein structures.



Map of the human X chromosome (from the NCBI website). Assembly of the human genome is one of the greatest achievements of bioinformatics.

The primary goal of bioinformatics is to increase our understanding of biological processes. What sets it apart from other approaches, however, is its focus on developing and applying computationally intensive techniques (e.g., data mining, and machine learning algorithms) to achieve this goal. Major research efforts in the field include sequence alignment, gene

finding, genome assembly, protein structure alignment, protein structure prediction, prediction of gene expression and protein-protein interactions, genome-wide association studies and the modeling of evolution.

Introduction

Bioinformatics was applied in the creation and maintenance of a database to store biological information at the beginning of the "genomic revolution", such as nucleotide and amino acid sequences. Development of this type of database involved not only design issues but the development of complex interfaces whereby researchers could both access existing data as well as submit new or revised data.

In order to study how normal cellular activities are altered in different disease states, the biological data must be combined to form a comprehensive picture of these activities. Therefore, the field of bioinformatics has evolved such that the most pressing task now involves the analysis and interpretation of various types of data, including nucleotide and amino acid sequences, protein domains, and protein structures. The actual process of analyzing and interpreting data is referred to as computational biology. Important sub-disciplines within bioinformatics and computational biology include:

a) the development and implementation of tools that enable efficient access to, and use and management of, various types of information. **b)** the development of new algorithms (mathematical formulas) and statistics with which to assess relationships among members of large data sets, such as methods to locate a gene within a sequence, predict protein structure and/or function, and cluster protein sequences into families of related sequences.

Major research areas

Sequence analysis

Since the Phage Φ-X174 was sequenced in 1977, the DNA sequences of hundreds of organisms have been decoded and stored in databases. The information is analyzed to determine genes that encode polypeptides, as well as regulatory sequences. A comparison of genes within a species or between different species can show similarities between protein functions, or relations between species (the use of molecular systematics to construct phylogenetic trees). With the growing amount of data, it long ago became impractical to analyze DNA sequences manually. Today, computer programs are used to search the genome of thousands of organisms, containing billions of nucleotides. These programs would compensate for mutations (exchanged, deleted or inserted bases) in the DNA sequence, in order to identify sequences that are related, but not identical. A variant of this sequence alignment is used in the sequencing process itself. The so-called shotgun sequencing technique (which was used, for example, by The Institute for Genomic Research to sequence the first bacterial genome, *Haemophilus influenzae*) does not give a sequential list of nucleotides, but instead the sequences of thousands of small DNA fragments (each about 600-800 nucleotides long). The ends of these fragments overlap and, when aligned in the right way, make up the complete genome. Shotgun sequencing yields sequence data quickly, but the task of assembling the fragments can be quite complicated for larger genomes. In the case of the Human Genome Project, it took several days of CPU time (on one hundred Pentium III desktop machines clustered specifically for the purpose) to assemble the fragments. Shotgun sequencing is the method of choice for virtually all

genomes sequenced today, and genome assembly algorithms are a critical area of bioinformatics research.

Another aspect of bioinformatics in sequence analysis is the automatic search for genes and regulatory sequences within a genome. Not all of the nucleotides within a genome are genes. Within the genome of higher organisms, large parts of the DNA do not serve any obvious purpose. This so-called junk DNA may, however, contain unrecognized functional elements. Bioinformatics helps to bridge the gap between genome and proteome projects--for example, in the use of DNA sequences for protein identification.

See also: sequence analysis, sequence profiling tool, sequence motif.

Genome annotation

In the context of → genomics, **annotation** is the process of marking the genes and other biological features in a DNA sequence. The first genome annotation software system was designed in 1995 by Dr. Owen White, who was part of the team that sequenced and analyzed the first genome of a free-living organism to be decoded, the bacterium *Haemophilus influenzae*. Dr. White built a software system to find the genes (places in the DNA sequence that encode a protein), the transfer RNA, and other features, and to make initial assignments of function to those genes. Most current genome annotation systems work similarly, but the programs available for analysis of genomic DNA are constantly changing and improving.

Computational evolutionary biology

Evolutionary biology is the study of the origin and descent of species, as well as their change over time. Informatics has assisted evolutionary biologists in several key ways; it has enabled researchers to:

- trace the evolution of a large number of organisms by measuring changes in their → DNA, rather than through physical taxonomy or physiological observations alone,
- more recently, compare entire genomes, which permits the study of more complex evolutionary events, such as gene duplication, horizontal gene transfer, and the prediction of factors important in bacterial speciation,
- build complex computational models of populations to predict the outcome of the system over time
- track and share information on an increasingly large number of species and organisms

Future work endeavours to reconstruct the now more complex tree of life.

The area of research within computer science that uses genetic algorithms is sometimes confused with computational evolutionary biology, but the two areas are unrelated.

Measuring biodiversity

Biodiversity of an ecosystem might be defined as the total genomic complement of a particular environment, from all of the species present, whether it is a biofilm in an abandoned mine, a drop of sea water, a scoop of soil, or the entire biosphere of the planet Earth. Databases are used to collect the species names, descriptions, distributions, genetic information, status and size of populations, habitat needs, and how each organism interacts with other species. Specialized software programs are used to find, visualize, and analyze the information, and most importantly, communicate it to other people. Computer

simulations model such things as population dynamics, or calculate the cumulative genetic health of a breeding pool (in agriculture) or endangered population (in conservation). One very exciting potential of this field is that entire → DNA sequences, or genomes of endangered species can be preserved, allowing the results of Nature's genetic experiment to be remembered *in silico*, and possibly reused in the future, even if that species is eventually lost.[1]

Analysis of gene expression

The expression of many genes can be determined by measuring mRNA levels with multiple techniques including microarrays, expressed cDNA sequence tag (EST) sequencing, serial analysis of gene expression (SAGE) tag sequencing, massively parallel signature sequencing (MPSS), or various applications of multiplexed in-situ hybridization. All of these techniques are extremely noise-prone and/or subject to bias in the biological measurement, and a major research area in computational biology involves developing statistical tools to separate signal from noise in high-throughput gene expression studies. Such studies are often used to determine the genes implicated in a disorder: one might compare microarray data from cancerous epithelial cells to data from non-cancerous cells to determine the transcripts that are up-regulated and down-regulated in a particular population of cancer cells.

Analysis of regulation

Regulation is the complex orchestration of events starting with an extracellular signal such as a hormone and leading to an increase or decrease in the activity of one or more proteins. Bioinformatics techniques have been applied to explore various steps in this process. For example, promoter analysis involves the identification and study of sequence motifs in the DNA surrounding the coding region of a gene. These motifs influence the extent to which that region is transcribed into mRNA. Expression data can be used to infer gene regulation: one might compare microarray data from a wide variety of states of an organism to form hypotheses about the genes involved in each state. In a single-cell organism, one might compare stages of the cell cycle, along with various stress conditions (heat shock, starvation, etc.). One can then apply clustering algorithms to that expression data to determine which genes are co-expressed. For example, the upstream regions (promoters) of co-expressed genes can be searched for over-represented regulatory elements.

Analysis of protein expression

Protein microarrays and high throughput (HT) mass spectrometry (MS) can provide a snapshot of the proteins present in a biological sample. Bioinformatics is very much involved in making sense of protein microarray and HT MS data; the former approach faces similar problems as with microarrays targeted at mRNA, the latter involves the problem of matching large amounts of mass data against predicted masses from protein sequence databases, and the complicated statistical analysis of samples where multiple, but incomplete peptides from each protein are detected.

Analysis of mutations in cancer

In cancer, the genomes of affected cells are rearranged in complex or even unpredictable ways. Massive sequencing efforts are used to identify previously unknown point mutations in a variety of genes in cancer. Bioinformaticians continue to produce specialized automated systems to manage the sheer volume of sequence data produced, and they create new algorithms and software to compare the sequencing results to the growing collection of human genome sequences and germline polymorphisms. New physical detection technology are employed, such as oligonucleotide microarrays to identify chromosomal gains and losses (called comparative genomic hybridization), and single nucleotide polymorphism arrays to detect known *point mutations*. These detection methods simultaneously measure several hundred thousand sites throughout the genome, and when used in high-throughput to measure thousands of samples, generate terabytes of data per experiment. Again the massive amounts and new types of data generate new opportunities for bioinformaticians. The data is often found to contain considerable variability, or noise, and thus Hidden Markov model and change-point analysis methods are being developed to infer real copy number changes.

Another type of data that requires novel informatics development is the analysis of lesions found to be recurrent among many tumors .

Prediction of protein structure

Protein structure prediction is another important application of bioinformatics. The amino acid sequence of a protein, the so-called primary structure, can be easily determined from the sequence on the gene that codes for it. In the vast majority of cases, this primary structure uniquely determines a structure in its native environment. (Of course, there are exceptions, such as the bovine spongiform encephalopathy - aka Mad Cow Disease - prion.) Knowledge of this structure is vital in understanding the function of the protein. For lack of better terms, structural information is usually classified as one of *secondary*, *tertiary* and *quaternary* structure. A viable general solution to such predictions remains an open problem. As of now, most efforts have been directed towards heuristics that work most of the time.

One of the key ideas in bioinformatics is the notion of homology. In the genomic branch of bioinformatics, homology is used to predict the function of a gene: if the sequence of gene *A*, whose function is known, is homologous to the sequence of gene *B,* whose function is unknown, one could infer that B may share A's function. In the structural branch of bioinformatics, homology is used to determine which parts of a protein are important in structure formation and interaction with other proteins. In a technique called homology modeling, this information is used to predict the structure of a protein once the structure of a homologous protein is known. This currently remains the only way to predict protein structures reliably.

One example of this is the similar protein homology between hemoglobin in humans and the hemoglobin in legumes (leghemoglobin). Both serve the same purpose of transporting oxygen in the organism. Though both of these proteins have completely different amino acid sequences, their protein structures are virtually identical, which reflects their near identical purposes.

Other techniques for predicting protein structure include protein threading and *de novo* (from scratch) physics-based modeling.

See also: structural motif and structural domain.

Comparative genomics

The core of comparative genome analysis is the establishment of the correspondence between genes (orthology analysis) or other genomic features in different organisms. It is these intergenomic maps that make it possible to trace the evolutionary processes responsible for the divergence of two genomes. A multitude of evolutionary events acting at various organizational levels shape genome evolution. At the lowest level, point mutations affect individual nucleotides. At a higher level, large chromosomal segments undergo duplication, lateral transfer, inversion, transposition, deletion and insertion. Ultimately, whole genomes are involved in processes of hybridization, polyploidization and endosymbiosis, often leading to rapid speciation. The complexity of genome evolution poses many exciting challenges to developers of mathematical models and algorithms, who have recourse to a spectra of algorithmic, statistical and mathematical techniques, ranging from exact, heuristics, fixed parameter and approximation algorithms for problems based on parsimony models to Markov Chain Monte Carlo algorithms for Bayesian analysis of problems based on probabilistic models.

Many of these studies are based on the homology detection and protein families computation.

Modeling biological systems

Systems biology involves the use of computer simulations of cellular subsystems (such as the → networks of metabolites and enzymes which comprise metabolism, signal transduction pathways and gene regulatory networks) to both analyze and visualize the complex connections of these cellular processes. Artificial life or virtual evolution attempts to understand evolutionary processes via the computer simulation of simple (artificial) life forms.

High-throughput image analysis

Computational technologies are used to accelerate or fully automate the processing, quantification and analysis of large amounts of high-information-content biomedical imagery. Modern image analysis systems augment an observer's ability to make measurements from a large or complex set of images, by improving accuracy, objectivity, or speed. A fully developed analysis system may completely replace the observer. Although these systems are not unique to biomedical imagery, biomedical imaging is becoming more important for both diagnostics and research. Some examples are:

- high-throughput and high-fidelity quantification and sub-cellular localization (high-content screening, cytohistopathology)
- morphometrics
- clinical image analysis and visualization
- determining the real-time air-flow patterns in breathing lungs of living animals
- quantifying occlusion size in real-time imagery from the development of and recovery during arterial injury
- making behavioral observations from extended video recordings of laboratory animals
- infrared measurements for metabolic activity determination
- inferring clone overlaps in DNA mapping, e.g. the Sulston score

Protein-protein docking

In the last two decades, tens of thousands of protein three-dimensional structures have been determined by X-ray crystallography and Protein nuclear magnetic resonance spectroscopy (protein NMR). One central question for the biological scientist is whether it is practical to predict possible protein-protein interactions only based on these 3D shapes, without doing → protein-protein interaction experiments. A variety of methods have been developed to tackle the Protein-protein docking problem, though it seems that there is still much work to be done in this field.

Software and tools

Software tools for bioinformatics range from simple command-line tools, to more complex graphical programs and standalone web-services available from various bioinformatics companies or public institutions. The computational biology tool best-known among biologists is probably BLAST, an algorithm for determining the similarity of arbitrary sequences against other sequences, possibly from curated databases of protein or DNA sequences. BLAST is one of a number of generally available programs for doing sequence alignment. The NCBI provides a popular web-based implementation that searches their databases.

Web services in bioinformatics

SOAP and REST-based interfaces have been developed for a wide variety of bioinformatics applications allowing an application running on one computer in one part of the world to use algorithms, data and computing resources on servers in other parts of the world. The main advantages lay in the end user not having to deal with software and database maintenance overheads. Basic bioinformatics services are classified by the EBI into three categories: SSS (Sequence Search Services), MSA (Multiple Sequence Alignment) and BSA (Biological Sequence Analysis). The availability of these service-oriented bioinformatics resources demonstrate the applicability of web based bioinformatics solutions, and range from a collection of standalone tools with a common data format under a single, standalone or web-based interface, to integrative, distributed and extensible bioinformatics workflow management systems.

See also

Related topics

- Biocybernetics
- Bioinformatics companies
- Biologically-inspired computing
- Biomedical informatics
- Computational biology
- Computational biomodeling
- Computational genomics
- DNA sequencing theory
- Dot plot (bioinformatics)
- Dry lab
- Margaret Oakley Dayhoff

- → Metabolic network modelling
- Molecular Design software
- Morphometrics
- Natural computation
- Pharmaceutical company
- Protein-protein interaction prediction
- List of nucleic acid simulation software
- List of numerical analysis software
- List of protein structure prediction software
- List of scientific journals in bioinformatics

Related fields

- Applied mathematics
- Artificial intelligence
- Biology
- Cheminformatics
- Clinomics
- Comparative genomics
- Computational biology
- Computational epigenetics
- Computational science
- Computer science
- Cybernetics
- Ecoinformatics
- → Genomics
- Informatics
- Information theory
- → Mathematical biology
- Molecular modelling
- Neuroinformatics
- → Proteomics
- Pervasive adaptation
- Scientific computing
- Statistics
- Structural biology
- → Systems biology
- Theoretical biology
- Veterinary informatics

References

[1] *Important projects:* Species 2000 project (http://www.sp2000.org/); uBio Project (http://www.ubio.org/); Partnership for Biodiversity Informatics (http://pbi.ecoinformatics.org/)

- Achuthsankar S Nair Computational Biology & Bioinformatics - A gentle Overview (http://print.achuth.googlepages.com/BINFTutorialV5.0CSI07.pdf), Communications of Computer Society of India, January 2007
- Aluru, Srinivas, ed. *Handbook of Computational Molecular Biology*. Chapman & Hall/Crc, 2006. ISBN 1584884061 (Chapman & Hall/Crc Computer and Information Science Series)
- Baldi, P and Brunak, S, *Bioinformatics: The Machine Learning Approach*, 2nd edition. MIT Press, 2001. ISBN 0-262-02506-X
- Barnes, M.R. and Gray, I.C., eds., *Bioinformatics for Geneticists*, first edition. Wiley, 2003. ISBN 0-470-84394-2
- Baxevanis, A.D. and Ouellette, B.F.F., eds., *Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins*, third edition. Wiley, 2005. ISBN 0-471-47878-4
- Baxevanis, A.D., Petsko, G.A., Stein, L.D., and Stormo, G.D., eds., *Current Protocols in Bioinformatics*. Wiley, 2007. ISBN 0-471-25093-7
- Claverie, J.M. and C. Notredame, *Bioinformatics for Dummies*. Wiley, 2003. ISBN 0-7645-1696-5
- Cristianini, N. and Hahn, M. *Introduction to Computational Genomics* (http://www.computational-genomics.net/), Cambridge University Press, 2006. (ISBN 9780521671910 | ISBN 0521671914)
- Durbin, R., S. Eddy, A. Krogh and G. Mitchison, *Biological sequence analysis*. Cambridge University Press, 1998. ISBN 0-521-62971-3
- Gilbert, D. *Bioinformatics software resources* (http://bib.oxfordjournals.org/cgi/content/abstract/5/3/300). Briefings in Bioinformatics, Briefings in Bioinformatics, 2004 5(3):300-304.
- Keedwell, E., *Intelligent Bioinformatics: The Application of Artificial Intelligence Techniques to Bioinformatics Problems*. Wiley, 2005. ISBN 0-470-02175-6
- Kohane, et al. *Microarrays for an Integrative Genomics*. The MIT Press, 2002. ISBN 0-262-11271-X
- Lund, O. et al. *Immunological Bioinformatics*. The MIT Press, 2005. ISBN 0-262-12280-4
- Michael S. Waterman, *Introduction to Computational Biology: Sequences, Maps and Genomes*. CRC Press, 1995. ISBN 0-412-99391-0
- Mount, David W. *Bioinformatics: Sequence and Genome Analysis* Spring Harbor Press, May 2002. ISBN 0-87969-608-7
- Pachter, Lior and Sturmfels, Bernd. "Algebraic Statistics for Computational Biology" Cambridge University Press, 2005. ISBN 0-521-85700-7
- Pevzner, Pavel A. *Computational Molecular Biology: An Algorithmic Approach* The MIT Press, 2000. ISBN 0-262-16197-4
- Tisdall, James. "Beginning Perl for Bioinformatics" O'Reilly, 2001. ISBN 0-596-00080-4
- Dedicated issue of *Philosophical Transactions B* on Bioinformatics freely available (http://publishing.royalsociety.org/bioinformatics)
- Catalyzing Inquiry at the Interface of Computing and Biology (2005) CSTB report (http://www.nap.edu/catalog/11480.html)
- Calculating the Secrets of Life: Contributions of the Mathematical Sciences and computing to Molecular Biology (1995) (http://www.nap.edu/catalog/2121.html)

- Foundations of Computational and Systems Biology MIT Course (http://ocw.mit.edu/OcwWeb/Biology/7-91JSpring2004/LectureNotes/index.htm)
- Computational Biology: Genomes, Networks, Evolution Free MIT Course (http://ocw.mit.edu/OcwWeb/Electrical-Engineering-and-Computer-Science/6-895Fall-2005/CourseHome/index.htm)
- Algorithms for Computational Biology Free MIT Course (http://ocw.mit.edu/OcwWeb/Electrical-Engineering-and-Computer-Science/6-096Spring-2005/CourseHome/index.htm)
- Zhang, Z., Cheung, K.H. and Townsend, J.P. Bringing Web 2.0 to bioinformatics, Briefing in Bioinformatics. In press (http://www.ncbi.nlm.nih.gov/pubmed/18842678)

External links

- Major Organizations
  - Bioinformatics Organization (Bioinformatics.Org): The Open-Access Institute (http://bioinformatics.org/)
  - EMBnet (http://www.embnet.org/)
  - European Bioinformatics Institute (http://www.ebi.ac.uk/)
  - European Molecular Biology Laboratory (http://www.embl.org/)
  - The International Society for Computational Biology (http://www.iscb.org/)
  - National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/)
  - National Institutes of Health homepage (http://www.nih.gov)
  - Open Bioinformatics Foundation: umbrella non-profit organization supporting certain open-source projects in bioinformatics (http://www.open-bio.org/)
  - Swiss Institute of Bioinformatics
  - Wellcome Trust Sanger Institute
- Major Journals
  - Algorithms in Molecular Biology (http://www.almob.org/)
  - Bioinformatics (http://bioinformatics.oupjournals.org/)
  - BMC Bioinformatics (http://www.biomedcentral.com/bmcbioinformatics)
  - Briefings in Bioinformatics (http://bib.oxfordjournals.org/)
  - Evolutionary Bioinformatics (http://www.la-press.com/evolbio.htm)
  - Genome Research (http://www.genome.org)
  - The International Journal of Biostatistics (http://www.bepress.com/ijb/)
  - Journal of Computational Biology (http://www.liebertpub.com/publication.aspx?pub_id=31)
  - Cancer Informatics (http://la-press.com/journal.php?pa=description&journal_id=10)
  - Journal of the Royal Society Interface (http://publishing.royalsociety.org/index.cfm?page=1058)
  - Molecular Systems Biology (http://www.nature.com/msb/index.html)
  - PLoS Computational Biology (http://compbiol.plosjournals.org)
  - Statistical Applications in Genetic and Molecular Biology (http://www.bepress.com/sagmb/)
  - Transactions on Computational Biology and Bioinformatics - IEEE/ACM (http://www.computer.org/tcbb/)
  - International Journal of Bioinformatics Research and Applications (http://www.inderscience.com/browse/index.php?journalcode=ijbra)

    - List of Bioinformatics journals (http://www.bioinformatics.fr/journals.php) at Bioinformatics.fr
    - EMBnet.News (http://www.embnet.org) at EMBnet.org
    - International Journal of Computational Biology and Drug Design (IJCBDD)
    - International Journal of Functional Informatics and Personalized Medicine (IJFIPM)
- Other sites
    - The Collection of Biostatistics Research Archive (http://www.biostatsresearch.com/repository/)
    - Human Genome Project and Bioinformatics (http://www.ornl.gov/TechResources/Human_Genome/research/informatics.html)
    - List of Bioinformatics Research Groups (http://www.bioinformatics.fr/laboratories.php) at Bioinformatics.fr
    - List of Bioinformatics Research Groups (http://www.dmoz.org/Science/Biology/Bioinformatics/Research_Groups//) at the Open Directory Project
- Tutorials / Resources / Primers
    - Bioinformatics - A Science Primer (http://www.ncbi.nlm.nih.gov/About/primer/bioinformatics.html) — by NCBI

See also

- International Society of Intelligent Biological Medicine (ISIBM)

Article Sources and Contributors

DNA *Source*: http://en.wikipedia.org/w/index.php?oldid=293218723 *Contributors*: (, (jarbarf), -Majestic-, 168.., 168..., 169, 17Drew, 3dscience, 4u1e, 62.253.64.xxx, 7434be, 84user, A D 13, A bit iffy, A-giau, Aaaxlp, Aatomic1, Academic Challenger, Acer, Adam Bishop, Adambiswanger1, Adamstevenson, Adashiel, Adenosine, Adrian.benko, Ahoerstemeier, Aitias, Aj123456, Alai, Alan Au, Aldaron, Aldie, Alegoo92, Alexandremas, Alkivar, Alphachimp, Alzhaid, Amboo85, Anarchy on DNA, Ancheta Wis, AndonicO, Andre Engels, Andrew wilson, Andreww, Andrij Kursetsky, Andycjp, Anita1988, Anomalocaris, Antandrus, Ante Aikio, Anthere, Anthony, Anthony Appleyard, Antilived, Antony-22, Aquaplus, Aquilla34, ArazZeynili, Arcadian, Ardyn, ArielGold, Armored Ear, Artichoker, Asbestos, Astrowob, Atlant, Aude, Autonova, Avala, AxelBoldt, AySz88, AzaToth, BD2412, BMF81, Banus, BaronLarf, Bbatsell, Bci2, Bcorr, Ben Webber, Ben-Zin, BenBildstein, Benjah-bmm27, Bensaccount, Bernie Sanders' DNA, Bevo, Bhadani, Bhar100101, BiH, Bijee, BikA06, Bill Nelson's DNA, Billmcgn189, Biolinker, Biriwilg, Bjwebb, Blastwizard, Blondtraillite, Bmtbomb, Bobblewik, Bobo192, Bongwarrior, Borisblue, Bornhj, Brian0918, Brighterorange, Brim, Brockett, Bryan, Bryan Derksen, CWY2190, Cacycle, Caerwine, Cainer91, Cal 1234, Calaschysm, Can't sleep, clown will eat me, Canadaduane, Carbon-16, Carcharoth, Carlo.milanesi, Carlwev, Casliber, Cathalgarvey, CatherineMunro, CattleGirl, Cburnett, Cerberus lord, Chanora, Chanting Fox, Charm, Chill Pill Bill, Chino, Chodges, Chris 73, Chris84, Chuck Grassley's DNA, Chuck02, Clivedelmonte, ClockworkSoul, CloudNine, Collins.mc, Colorajo, CommonsDelinker, Conversion script, Cool3, Coolawesome, Coredesat, Cornacchia123, Cosmotron, Cradlelover123, Crazycomputers, Crowstar, Crusadeonilliteracy, CryptoDerk, Crzrussian, Cubskrazy29, CupOBeans, Curps, Cyan, Cyclonenim, Cyrius, D6, DIREKTOR, DJAX, DJRafe, DNA EDIT WAR, DNA is shyt, DVD R W, Daniel Olsen, Daniel987600, Danielkueh, Danny, Danny B-), Danski14, Darklilac, Darth Panda, Davegrupp, David D., David Eppstein, Davidbspalding, Daycd, Db099221, Dbabbitt, Dcoetzee, DeAceShooter, DeadEyeArrow, Delldot, Delta G, Deltabeignet, DevastatorIIC, Diberri, Dicklyon, Digger3000, Digitalme, Dina, Djm1279, Dlohcierekim's sock, Dmn, Docjames, Doctor Faust, Docu, DonSiano, Donarreiskoffer, Dr d12, Dr.Kerr, Drini, Dudewheresmywallet, Dullhunk, Duncan.france, Dungodung, Dysmorodrepanis, E. Wayne, ERcheck, ESkog, Echo park00, Echuck215, Eddycrai, Editing DNA, Edwy, Efbfweborg, Egil, ElTyrant, Elb2000, Eleassar777, EliasAlucard, ElinorD, Ellmist, Eloquence, Emoticon, Epingchris, Erik Zachte, Escape Artist Swyer, Esurnir, Etanol, Ettrig, EurekaLott, Everyking, Evil Monkey, Ewawer, Execvator, FOTEMEH, Fabhcún, Factual, Fagstein, Fastfission, Fconaway, Fcrick, Fernando S. Aldado, Ffirehorse, Figma, Figure, Firefoxman, Firetrap9254, Fishingpal99, Flavaflav1005, Florentino floro, Fnielsen, Forluvoft, Freakofnurture, FreplySpang, Friendly Neighbour, Frostyservant, Fruge, Fvasconcellos, G3pro, GAThrawn22, GHe, GODhack, Gaara san, Galoubet, Gary King, Gatortpk, Gazibara, GeeJo, Gene Nygaard, GeoMor, Giftlite, Gilisa, Gilliam, Gimmetrow, Gjuggler, Glen Hunt's DNA, Glenn, Gmaxwell, GoEThe, Goatasaur, Gogo Dodo, Golnazfotohabadi, GordonWatts, Gracenotes, Graeme Bartlett, GraemeL, Grafikm fr, Graft, Graham87, GrahamColm, Grandegrandegrande, GregorB, Grover cleveland, Gurko, Gustav von Humpelschmumpel, Gutza, Gwsrinme, Hadal, Hagerman, Hairchrm, Hairwheel, Hannes Röst, Harianto, Heathhunnicutt, Hephaestos, Heron, Heyheyhack, Hockey21dude, Horatio, Hu, Hughdbrown, Hurricanehink, Hut 8.5, Hvn0413, I Seek To Help & Repair!, I hate DNA, Iapetus, Icairns, Ilia Kr., Impamiizgraa, InShaneee, Inge-Lyubov, Isilanes, Isis07, Itub, Ixfd64, Izehar, JHMM13, JWSchmidt, JWSurf, Jacek Kendysz, Jackrm, JamesMLane, JamesMt1984, Janejellyroll, Jaxl, Jeka911, Jer ome, JeremyA, Jerzy, Jetsetpainter, Jh51681, Jiddisch, Jimriz, Jimwong, Jlh29, Jls043, Jmcc150, Jo9100, JoanneB, Joconnol, Johanvs, JohnArmagh, Johntex, Johnuniq, Jojit fb, JonMoulton, Jonrunles, Jorvik, JoshuaZ, Josq, Jossi, Jstech, Julian Diamond, Jumbo Snails, Junes, Jwrosenzweig, Kahlfin, Kapow, Karrmann, Kazkaskazkasako, Kbh3rd, Keegan, Keepweek, Keilana, Kelly Martin, Kemyou, Kendrick7, KerryO77, Kevin Breitenstein, Kevmitch, Kghose, Kholdstare99, Kierano, KimvdLinde, King of Hearts, KingTT, Kingturtle, Kitch, Knaggs, Knowledge Seeker, KnowledgeOfSelf, Koavf, KrakatoaKatie, Kums, Kungfuadam, Kuru, Kwamikagami, Kwekubo, KyNephi, LA2, La goutte de pluie, Lascorz, Latka, Lavateraguy, Lee Daniel Crocker, Lemchesvej, Lerdsuwa, Leuko, Lexor, Lhenslee, Lia Todua, LightFlare, Lightmouse, Lightspeedchick, Ligulem, Lincher, Lion Wilson, Lir, Llongland, Llull, Lockesdonkey, Logical2u, Loginbuddy, Looxix, Loren36, Loris, Luigi30, Luk, Lumos3, Luna Santin, Luuva, MER-C, MKoltnow, MONGO, Mac, Madeleine Price Ball, Madhero88, Magadan, Magnus Manske, Majorly, Malcolm rowe, Malo, Mandyj61596, Mantissa128, Marcus.aerlous, Marj Tiefert, MarvPaule, Master dingley, Mattbr, Mattbrundage, Mattjblythe, Mav, Max Baucus' DNA, Max Naylor, McDogm, Medessec, Medos2, Melaen, Melchoir, Mentalmaniac07, Mgiganteus1, Mgtoohey, Mhking, Michael Devore, MichaelHa, Michaelas10, Michigan user, MidgleyDJ, Midnightblueowl, Midoriko, Mika293, Mike Rosoft, Mikker, Mikko Paananen, Mintman16, MisfitToys, Misza13, Mithent, Mjpieters, Mleefs7, Moink, Moorice, Mortene, Mr Bungle, Mr Meow Meow, Mr Stephen, MrErku, Mstislavl, Mstroeck, Mulad, Munita Prasad, Muro de Aguas, Mwanner, Mxn, Nakon, Narayanese, Natalie Erin, Natarajanganesan, Nate1028, NatureA16, Nauseam, Nbauman, Neckro, Netkinetic, Netoholic, Neutrality, NewEnglandYankee, Nighthawk380, NighthawkJ, Nihiltres, Nirajrm, Nishkid64, Nitecrawler, Nitramrekcap, No Guru, NoIdeaNick, NochnoiDozor, Nohat, Northfox, NorwegianBlue, Nthornberry, Nunh-huh, OBloodyHell, OOODDD, Obli, Oblivious, Ocolon, Ojl, Omicronpersei8, Onco p53, Opabinia regalis, Opelio, Orrin Hatch's DNA, Orthologist, Ortolan88, Ouishoebean, Outriggr, OwenX, P99am, PDH, PFHLai, PaePae, Pakaran, Pascal666, Patrick0Moran, Patrick2480, Patstuart, Paul venter, Paulinho28, Pcb21, Pde, Peak, Pedro, Persian Poet Gal, Peter Isotalo, Peter K., Peter Winnberg, Pgan002, Philip Trueman, PhilipO, Phoenix Hacker, Pierceno, PierreAbbat, Pigman, Pigmietheclub, Pilotguy, Pkirlin, Poor Yorick, Portugue6927, Potatoswatter, Preston47, Priscilla 95925, Pristontale111, Pro crast in a tor, Prodego, Psora, PsyMar, Psymier, Pumpkingrower05, Pyrospirit, Quebec99, Quickbeam, Qutezuce, Qxz, R'n'B, R. S. Shaw, RDBrown, RSido, Ragesoss, Rajwiki123, RandomP, Randomblue, Raul654, Raven in Orbit, Ravidreams, Rdb, Rdsmith4, Red Director, Reddi, Rednblu, Redneckjimmy, Redquark, Retired username, Rettetast, RexNL, Rich Farmbrough, RichG, Richard Durbin's DNA, Ricky81682, Rjwilmsi, Roadnottaken, Robdurbar, RobertG, Rocastelo, RoddyYoung, Rory096, Rotem Dan, Roy Brumback, RoyBoy, RoyLaurie, Royalguard11, RunOrDie, Russ47025, RxS, Ryan Delaney, RyanGerbil10, Ryulong, S77914767, SCEhardt, STAN SWANSON, SWAdair, Sabbre, Safwan40, Sakkura, SallyForth123, Sam Burne James, Samsara, Samuel, Samuel Blanning, SandyGeorgia, Sango123, Sangwine, Savidan, Sceptre, Schutz, Sciencechick, Scienceman123, Scincesociety, Sciurinæ, Scope creep, Scoterican, Sean William, SeanMack, Seans Potato Business, SebastianHawes, Seldon1, Serephine, Shadowlynk, Shanes, ShaunL, Shekharsuman, Shizhao, Shmee47, Shoy, Silsor, SimonD, Sintaku, Sir.Loin, Sjjupadhyay, Sjollema, Sloth monkey, Slrubenstein, Sly G, SmilesALot, Smithbrenon, Snowmanradio, Snowolf, Snurks, Solipsist, Someone else, Sonett72, Sopoforic, Spaully, Spectrogram, Splette, Spondoolicks, Spongebobsqpants, SpuriousQ, Squidonius, Squirepants101, Statsone, Steel, Steinsky, Stemonitis, Stephenb, SteveHopson, Stevertigo, Stevietheman, Stewartadcock, Stuart7m, Stuhacking, SupaStarGirl, Supspirit, Susvolans, Sverdrup, Swid, Switchercat, T'Shael, Taco325i, Takometer, TakuyaMurata, Tariqabjotu, Tarret, Taulant23, Tavilis, Tazmaniacs, Ted Longstaffe, Tellyaddict, TenOfAllTrades, Terraguy, Test100000, TestPilot, The Rambling Man, The WikiWhippet, TheAlphaWolf, TheChrisD, TheGrza, TheKMan, TheRanger, Thorwald, ThreeDaysGraceFan101, Thue, Tiddly Tom, Tide rolls, TigerShark, TimVickers, Timewatcher, Timir2, Timl2k4, Timrollpickering, Timwi, Tobogganoggin, Toby Bartels, TobyWilson1992, Tom Allen, Tom Harkin's DNA, Tomgally, Toninu, Tony1, Tonyrenploki, Trd300gt, Trent Lott's DNA, Triwbe, Troels Arvin, Tstrobaugh, Tufflaw, Turnstep, Twilight Realm, Ty146Ty146, UBeR, Unint, Unukorno, Usergreatpower, Utcursch, Uthbrian, Vaernnond, Vandelizer, Vanished user, Vary, Virtualphtn, Visium, Vividonset, VladimirKorablin, Vsmith, Vyasa, WAS 4.250, WAvegetarian, WHeimbigner, WJBscribe, Wafulz, WarthogDemon, Wavelength, WelshMatt, West Brom 4ever, Where, Whosasking, Whoutz, Why My Fleece?, Wik, Wiki alf, Wiki emma johnson, Wikiborg, Wikipedia Administration, William Pietri, WillowW, Wimt, Wknight94, Wmahan, Wnt, Wobble, WolfmanSF, Wouterstomp, Wwwwolf, Xy7, YOUR DNA, Yahel Guhan, Yamamoto Ichiro, Yamla, YanWong, Yansa, Yaser al-Nabriss, Yasha, Yomama9753, Younusporteous, Yurik, ZScout370, Zahid Abdassabur, Zahiri, Zazou, Zell Miller's DNA, Zephyris, Zoicon5, Zouavman Le Zouave, Zsinj, Zven, 1329 anonymous edits

Molecular models of DNA *Source*: http://en.wikipedia.org/w/index.php?oldid=292162887 *Contributors*: Bci2, Chris the speller, Oscarthecat

Genomics *Source*: http://en.wikipedia.org/w/index.php?oldid=287806168 *Contributors*: *drew, 5dPZ, AdamRetchless, Adenosine, Alex naish, Andreadb, Anthere, Apers0n, Aphextwin5678, AxelBoldt, Barrylb, Bill.albing, Braidwood, Branttudor, Brion VIBBER, Bryan Derksen, Calvinthe1337, Ceyockey, Combio, CommodiCast, DabMachine, Dave Nelson, David D., Dekisugi, Dicklyon, Dmb000006, DoctorDNA, Dolfin, Drgarden, El C, Eubulides, Eugene, Fred Bradstadt, Gary King, Genometer, GeoMor, Ghostoroy, Giftlite, Gilliam, Habj, Hadal, Hbent, Heron, Jenks, Jethero, Jfdwolff, Joconnol, Joerg Kurt Wegner, Johntex, Johnuniq, Jongbhak, Larssono, Lexor, Lightmouse, Lost-theory, Mariusz Biegacki, Marj Tiefert, Mav, Mike Lin, Natarajanganesan, Nitwitpicker, Oleginger, Para, Peak, Pgan002, Pharmtao, Pion, Pvosta, Quizkajer, RandomP, Recury, Rein0299, Rich Farmbrough, Ronz, Rppgen, Sairen42, Scewing, Shanes, SimonP, Sjjupadhyay, Spitfire ch, Springmn, Starshadow, Stonedhamlet, Syp, Template namespace initialisation script, TheObtuseAngleOfDoom, Thkim75, Thorwald, Tiddly Tom, Toddst1, Touchstone42, Unyoyega, VashiDonsk, W09110900, Wavelength, Wayne530, Williamb, Wmahan, Wuzzybaba, Xanthoptica, ZayZayEM, ZimZalaBim, 128 anonymous edits

Proteomics *Source*: http://en.wikipedia.org/w/index.php?oldid=290412197 *Contributors*: 2over0, Aiko, Akriasas, Alan Liefting, AlistairMcMillan, Apfelsine, ArazZeynili, Babbage, Bdekker, Bezapt, Bill.albing, Borgx, Boy in the bands, Bryan Derksen, Calimo, CathCarey, Chaos, Chris the speller, Cjb88, Clicketyclack, Cpiggee, Dancter, Dave Nelson, Dfornika, Dhart, Dicklyon, Djstates, Dmb000006, Download, El C, Flowanda, Gacggt, Gaius Cornelius, GraemeLeggett, Graham87, Hadal, Iamunknown, IlyaHaykinson, Iridescent, Itub, Jambell, Janbrogger, Jason.nunes, JeLuF, Jfdwolff, Johannesvillavelius, JonHarder, Jóna Þórunn, Kbelhajj, Kevyn, Kjaergaard, Kkmurray, Kku, Kosigrim, Kukini, Lexor, LiDaobing, Lights, Lupin, Lysdexia, MStreble, Maartenvdv, Mani1, Manyanswer, Mathisuresh, Mav, Mjensen@nas.edu, N2e, Nick Y., Nina Gerlach, Nwbeeson, Oddwick, Oleginger, Ottava Rima, PDH, Paul Drye, Pcarvalho, Perissinotti, Pgan002, Pganas, Plumbago, Proteomicon, ProveIt, Pscott22, Pvosta, Quintote, RDBrown, Raymond Hui,

Remi0o, Rich Farmbrough, Roadnottaken, Sater, Schutz, Senski, Shizhao, Smcarlson, Someguy1221, Sorfane, Sp3000, Springatlast, Srlasky, StevieNic, Systemfolder, Template namespace initialisation script, TestPilot, Tim@, Tregonsee, Trevor MacInnis, Triwbe, Tstrobaugh, Versus22, Voyagerfan5761, Whosasking, Wisdom89, Xeaa, Zashaw, ZimZalaBim, आशीष भटनागर, 211 anonymous edits

Protein-protein interaction *Source*: http://en.wikipedia.org/w/index.php?oldid=293344844 *Contributors*: 56869kltaylor, 7bd1, A wandering 1, Alboyle, Apfelsine, Ashcroft, Bci2, Biophys, Clicketyclack, Cpichardo, D-rew, DarkSaber2k, Delldot, Djstates, Dsome, FreeKill, Giftlite, GracelinTina, Hendrik Fuß, Hotheartdog, Jeandré du Toit, Jkbioinfo, Jkwaran, Jn3v16, Jongbhak, Keesiewonder, Kkmurray, Kuheli, Kyawtun, Lafw, Lemchesvej, Lenticel, Longhair, Meb025, Michael Hardy, MichaelMcGuffin, Miguel Andrade, NickelShoe, Ninjagecko, Nnh, Rajah, Reb42, Riana, Ronz, Seans Potato Business, Snowolf, TheParanoidOne, Thorwald, Uthbrian, Victor D, Wenzelr, Whosasking, Wintrag, 66 anonymous edits

Metabolic network *Source*: http://en.wikipedia.org/w/index.php?oldid=231870375 *Contributors*: Blastwizard, Ceyockey, Oleginger, PDH, TheParanoidOne, TimVickers, Zephyris, 2 anonymous edits

Metabolic network modelling *Source*: http://en.wikipedia.org/w/index.php?oldid=285580489 *Contributors*: AKC, Barticus88, Choster, Dylan Lake, Glane23, Karthik.raman, Leptictidium, Mdd, Noalaignorancia, PDH, Parakkum, Piotrus, Ragesoss, Sarebi, TimVickers, Uncle G, 20 anonymous edits

Metabolic pathway *Source*: http://en.wikipedia.org/w/index.php?oldid=281327489 *Contributors*: Ahoerstemeier, Albert.so, Arcadian, Barticus88, Bensaccount, Bobo192, Brendanliamboyle, Centrx, Ceyockey, Clicketyclack, Conversion script, David D., Daylite, Delta G, Dposse, Dreg743, Drphilharmonic, Fenteany, GraemeL, Gregogil, HorsePunchKid, Jag123, Jamesters, Jfdwolff, Kku, Klunz, La goutte de pluie, Leptictidium, Lexor, Marshman, Modify, Monkeyflower, Muad, Parakkum, Pgan002, Qinatan, Raz1el, Ronz, RyanGerbil10, Sameetmehta, Shrimp wong, Sir marek, Snobscure, Stevietheman, TUF-KAT, Thebeginning, ThinkerThoughts, TimVickers, Tkynerd, Ukexpat, Unfree, Until It Sleeps, Whosyourjudas, Wrathchild, Zephyris, Zoicon5, 72 anonymous edits

Interaction network *Source*: http://en.wikipedia.org/w/index.php?oldid=289976450 *Contributors*: Abduallah mohammed, Jongbhak, Pichpich, Ronz, 3 anonymous edits

Interactomics *Source*: http://en.wikipedia.org/w/index.php?oldid=293358442 *Contributors*: Bci2, Bdevrees, Erick.Antezana, Erodium, Jong, Jongbhak, Karthik.raman, Lexor, Llull, Niteowlneils, PDH, Pekaje, Rajah, Tucsontt, 8 anonymous edits

Mathematical biology *Source*: http://en.wikipedia.org/w/index.php?oldid=293117766 *Contributors*: Adoniscik, Agilemolecule, Agricola44, Alan Liefting, Anclation, Andreas td, Aua, Audriusa, Bci2, Bduke, Berland, BillWSmithJr, Ceyockey, Charvest, Chopchopwhitey, Commander Nemet, Constructive editor, Cquan, Den fjättrade ankan, Durova, Dysprosia, Eduardoporcher, Fredrik, Gandalfxviv, Geronimo20, Guettarda, Henriok, Honeydew, Imoen, Jag123, Jaibe, Jennavecia, JonHarder, Jonsafari, Jpbowen, Jwdietrich2, Karl-Henner, Kripkenstein, Leptictidium, Lexor, Lquilter, M stone, MATThematical, Malcolm Farmer, Mathmoclaire, Maurreen, Melcombe, Michael Hardy, Oldekop, Oli Filth, Open4D, Owlmonkey, Percy Snoodle, PeterStJohn, PhDP, Plw, Porcher, Rich Farmbrough, Sintaku, Sir hubert, Squidonius, Ssavelan, StN, Stemonitis, Tompw, Triwbe, Vina, Wavelength, 68 anonymous edits

Systems biology *Source*: http://en.wikipedia.org/w/index.php?oldid=292526702 *Contributors*: APH, Aciel, Alan Liefting, AlirezaShaneh, Amandadawnbesemer, Amirsnik, Andreas td, Arthena, Asadrahman, Aua, Bad Cat, Bamess, BatteryIncluded, Bci2, Bio-ITWorld, Biochaos, Biophysik, Blueleezard, Boku wa kage, Broadbeer, CRGreathouse, CX, Can't sleep, clown will eat me, Captain-tucker, Ceolas, Charlenelieu, CharonZ, Ckatz, Claronow, ColinGillespie, Cquan, Crodriguel, D6, DanielNuyu, Delta Xi, Dmb000006, Drgarden, Droyarzun, Duelist135, Edaddison, Edward, Electric sheep, Erick.Antezana, Erkan Yilmaz, Eveillar, FLeader, Fences and windows, Fenice, Fletcher04, Foggy29, Fredrik, Garychurchill, Gcm, Gdrahnier, Ggonnell, Giftlite, Gwolfe, Heisner, HexiToFor, IPSOS, JRSocInterface, JaGa, Jdegreef, Jethero, Jondel, Jongbhak, Jpbowen, JulioVeraGon, Jwdietrich2, Kane5187, Karthik.raman, Kcordina, KirbyRandolf, Kku, Klenod, Klipkow, Lauranrg, Lenov, Letranova, Lexor, Lilia Alberghina, Linkman21, Lkathmann, Massbiotech, Mdd, Michael Fourman, Michael Hardy, Miguel Andrade, MikeHucka, Mkotl, Mkuiper, Mmxx, Mobashirgenome, Molelect, N2e, NBeale, NIH Media, Natelewis, Nbaliga, Neilbeach, Nemenman, Netsnipe, Nick Green, O RLY?, Ombudsman, Opertinicy, Patho, PaulGarner, Pkahlem, Pvosta, Quarl, Rajah, Reggiebird, Rich Farmbrough, Rjwilmsi, Robnpov, Rvencio, Rwcitek, Satish.vammi, SeeGee, Senu, Seuss01, Sholto Maud, Srlasky, Steinsky, Stewartadcock, Synthetic Biologist, Tagishsimon, Template namespace initialisation script, Thorwald, Triamus, Unauthorised Immunophysicist, Vangos, Versus22, Vonkje, WLU, Waltpohl, Wavelength, Whosasking, Xeaa, Zargulon, Zlite, Zoicon5, عیسی, 350 anonymous edits

Biotechnology *Source*: http://en.wikipedia.org/w/index.php?oldid=293230861 *Contributors*: -Majestic-, 209.144.103.xxx, 2D, AAMiller, ABF, APH, Abu al neez, Academic Challenger, Adenosine, Admiral Roo, Ageekgal, Aguerrap, Ahoerstemeier, Akita86, Akshaymat, Al Lemos, Alan Liefting, Alex.muller, Alex122188, Alexius08, Alfirin, AlistairMcMillan, Allstarecho, Andre Engels, Andycjp, Angela, Animum, Ankur Banerjee, Antandrus, Apoc2400, ArglebargleIV, ArielGold, Artaxiad, Arzachel, Baxter9, Bbbbbbbbbb, Beetstra, Belligero, Berkunt, Bhadani, Binscent, Biovini, Bjcairns, Blanchardb, BlaseDavid, Bleash, Bmeguru, Bob, Bobblewik, Bobo192, Bon d'une cythare, Bongwarrior, Bornhj, Brianga, Brusegadi, Bryan Derksen, Brynn Fulghum, Btan2525, Burto88, CIreland, CaSarge, Camembert, CanadianLinuxUser, Canis Lupus, CanisRufus, Canterbury Tail, Carinemily, Cessator, Ceyockey, Chaos, Che829, Chmod007, Chowbok, ChrisIsBelow, Chriso14, Cilynx, Cinnamon colbert, Ckatz, Clark89, ClockworkSoul, Closedmouth, Cnu.6914, Cnu4196, Cometstyles, CommodiCast, Computerjoe, Conversion script, Coolcaesar, Corpx, Cquan, Crimson Observer, Crusadeonilliteracy, Ctroy36, Curps, Cyfal, D, DIG, DXRAW, Da monster under your bed, Dbfirs, DeadEyeArrow, DerHexer, Derek.cashman, Discospinster, Dlae, DocWatson42, Douglas R. White, DougsTech, Doulos Christos, Drkarthi, Drumbeatsofeden, Dudesleeper, ERcheck, Ebrooks, Echosmoke, Economyweb, Edcolins, Editore99, Edward, Elbperle, Elpom, Emerydora, Engineer2020, Ennadaiit, Epbr123, Eric-Wester, Erikbinder, Eubulides, Eve Hall, Eviannab, FDP1, Fairy12, Fibonacci, Fjmustak, Flaviolin, Flying Jazz, Franamax, Fre5678, Frecklefoot, FreeKill, FreplySpang, Freyr, GJeffery, Gabriel Kielland, Gaius Cornelius, Galib20, Galoubet, Garion96, Gary King, Ghewgill, Giftlite, Gilliam, Glaurie, Gracenotes, Graham87, Ground Zero, Gurch, Happymercury, Harvardguy, Hdt83, Herd of Swine, HexaChord, Hqb, Hu12, Hughdbrown, Hulagutten, Hulek, Hunt 4 Orange November, Husond, I-think, Iant, Ibpassociation, IceUnshattered, Icey, Igoldste, Ike9898, Ikh, Ill Dilettante, ImperfectlyInformed, Indon, Intranetusa, Isporter, J.delanoy, JBKramer, JForget, JRR Trollkien, JWSchmidt, Jab843, Jack-A-Roe, Jaimetex, James086, Jamessuffield, Jason One, Jebba, Jensbn, Jerryseinfeld, JoanneB, John254, JohnCD, Johnfravolda, Johntex, JonHarder, Jossi, Jreconomy, Juliancolton, Kakofonous, Kandar, KaragouniS, Kbdank71, Keesiewonder, Keicia, Kharyal, Kic 423, Kimiko, Kingpin13, Kirthik.rugs, Kku, Kmtrapp, Knowledge Seeker, KnowledgeOfSelf, Kodiak71, Koolguyrock, Kpjas, Ktthebird, Kukini, Kungming2, Kuru, LAX, La Parka Your Car, La Pianista, Lamro, Lavish.aggarwal, Lawrence Cohen, Laxori666, LeaveSleaves, Lelkesa, Lendorien, Lexor, Leytonwd, Lginger, LiDaobing, Lickamaloin, Lightmouse, Lights, Ligulem, Lilac Soul, Linkspamremover, Llull, Looper5920, Lowellian, Lunboks, Lyellin, M don15, MER-C, MPerel, MTA2007, Mac, Mackolich, Madadem, Madhero88, Magnus Manske, Mahanga, Maphyche, Marek69, MarkSutton, Markuslarsson, MarleneG, Martin451, MartinDK, Martindo, Massbiotech, Master of Puppets, Matt poo watt, Maurreen, Maximilli, Mburbano, Mdwyer, Menchi, Michael Devore, MichaelJanich, Miguel in Portugal, Mild Bill Hiccup, Mishuletz, Misza13, Moondyne, Mrs Trellis, Muriel Gottrop, Mxn, Narayanese, NawlinWiki, Ndenison, Nectarflowed, NewEnglandYankee, Nicholas Cimini, O, OcciMoron, Ocee, OldakQuill, Oleg Alexandrov, Oli Filth, On the other side, Onco p53, Ossmann, OverlordQ, Oxymoron83, PDH, PGWG, Panarjedde, ParisianBlade, PeaceNT, Peak, Pearle, Petri Krohn, Philip Trueman, Pinkadelica, Prasanth.palukuri, Prodego, Pur2hit, Pwhited39, Pyrospirit, Qazwsxedcpolkm, Qtac, Quadell, QuadrivialMind, RHaworth, RONALD GARNER, RWFanMS, Randommouse, Ranveig, RazorICE, RedWolf, Remi0o, Rholton, Riana, Rich Farmbrough, Rifleman 82, Rkaufman13, Rkilman, Rmzelle, Rob Hooft, Rockvee, Rodak1, Ryan Delaney, RyanCross, S.K., SMC89, Sabisteb, SableSynthesis, Sambarwell, SamuelTheGhost, ScAvenger, SchfiftyThree, SchmuckyTheCat, Schzmo, Scientizzle, ScottJ, Sean D Martin, Seba5618, Seidenstud, Senator Palpatine, Serinde, Shaggorama, Shekharsuman, Showauthor, Shukig, Shuochli, SilkTork, Silversink, SimonP, Siriusfarm, Sjjupadhyay, Slakr, Sligocki, Smeira, Snowolf, Somearemoreequal, SpLoT, Spartan-James, Srleffler, Sstrumello, StAkAr Karnak, StarTrekkie, Steel, Stemonitis, StephanieM, Storm Rider, Styrofoam1994, Suisui, Talon Artaine, Tccarmichael, Telcourbanio, Template namespace initialisation script, Tempodivalse, TestPilot, TheNewEvolver, Thehelpfulone, Thingg, Thumar, Tidus773, TimVickers, Tirronan, Tom harrison, Tomchiukc, Tompagenet, Tomwac, Touchstone42, Traroth, Tregoweth, Tsmithnal, Tunatuna, Tutmosis, UberScienceNerd, Ukexpat, Ukt-zero, Until It Sleeps, User2004, Vanderesch, Vanka5, Ve2jgs, VengeancePrime, Vespristiano, Victor D, Vikrammat, Vlmastra, Vsmith, Walden, Wallet, Wapcaplet, WatermelonPotion, Wavelength, Werdna, White Cat, WikHead, Wiki alf, Wikiborg, Will Beback, William Avery, Wmahan, WriterHound, X!, X201, XP1, Xavexgoem, Xompanthy, Yosri, Zacheikov, Zereshk, Zurishaddai, Zzuuzz, 1122 anonymous edits

Bioinformatics *Source*: http://en.wikipedia.org/w/index.php?oldid=293336433 *Contributors*: 168..., 16@r, 3mta3, APH, Acerperi, Adenosine, Aetkin, Agricola44, AhmedMoustafa, Ahoerstemeier, Ajkarloss, Akpakp, Akriasas, Alai, Alan Au, Alex Kosorukoff, Amandadawnbesemer, Ambertk, Andersduck, Andkaha, Andreas C, AndriuZ, Angelsh, Ansell, ArglebargleIV, Artgen, Asasia, Ashalatha.jangala, Ashcroft, Asidhu, AuGold, Avenue, Azazello, Bact, Badanedwa, Banazir, Banus, Barticus88, Bcheng23, Bill.albing, Bill37212, Bio-ITWorld, Bioinformaticsguru, Bioinformin, Biovini, Blastwizard, Bm

richard, Bmeguru, Bmunro, Bob, Bobblewik, Bonnarj, Bonus Onus, Bookandcoffee, Bornslippy, Bradenripple, Brona, Burningsquid, Can't sleep, clown will eat me, Carey Evans, Cavrdg, Cbergman, Cbock, Chameleon, Chasingsol, Cholling, Chopchopwhitey, Christopherlin, Colin gravill, Colonialdirt, CommodiCast, ConceptExp, Conversion script, Counsell, Cquan, CryptoDerk, Cyc, Cyde, DIG, Dan198792, Dave Messina, David Ardell, David Gerard, Dismas, Dmb000006, Dod1, Don G., DonSiano, Donarreiskoffer, Dr02115, Dtabb, Dullhunk, Dysprosia, EALacey, EdGl, EdJohnston, Edgar181, Edward, Efbfweborg, Ehheh, El C, Ensignyu, Epbr123, Eramesan, Fcrozat, FireBrandon, Foscoe, Fotinakis, Frap, FreeKill, G716, GLHamilton, Ganeshbio1, Gaurav, Gazpacho, Gene s, Genometer, Giftlite, Girlwithglasses, Glen, Gonfus, Googed, Gordon014, GraemeL, Gulan722, Hawksj, HenkvD, Henriettaminge, Hike395, Hillarivallen, HoopyFrood, Imjustmatthew, Iwaterpolo, JHunterJ, Jamelan, Jameslyonsweiler, Jamiejoseph, Jchusid, Jcuticchia, Jengeldk, Jethero, Jimmaths, Jjwilkerson, Jkbioinfo, Joconnol, Joelrex, Joeoettinger, Joerg Kurt Wegner, JonHarder, Jorfer, JosephBarillari, Josephholsten, Joychen2010, Kamleong, Karol Langner, Keesiewonder, Kevin Breitenstein, Kevin.cohen, Kiwi2795, Kkmurray, Kku, Kotsiantis, Larry laptop, LeeWatts, Leofer, Lexor, Littlealien182, MER-C, MacGyverMagic, Macha, Madeleine Price Ball, Malafaya, Malcolm Farmer, Malkinann, Marashie, Marcoacostareyes, Martin Jambon, Martin.jambon, Mateo LeFou, MattWBradbury, Mattigatti, Mav, Mayumashu, Mazi, Mbadri, Metahacker, Michael Hardy, MicroBio Hawk, Mike Yang, Mindmatrix, Minho Bio Lee, Minimice, Mobashirgenome, Mstrangwick, Muchness, Muijz, Mxn, My walker 88, Nabeelbasheer, Natalya, Natarajanganesan, Navigatorwiki, Neksa, Nervexmachina, Nihiltres, Nivix, Ohnoitsjamie, Oleg Alexandrov, Oleginger, Opabinia regalis, Otets, P99am, PDH, PJY, Parakkum, Pascal.hingamp, Pawyilee, Pde, Peak, Perada, Perfectlover, Peter Znamenskiy, Phismith, Piano non troppo, Porcher, Postdoc, Ppgardne, Praveen pillay, Protonk, Pselvakumar, Pseudomonas, Quadell, Qwertyus, Raul654, Redgecko, Reinyday, Remi0o, Renji143, Rgonzaga, Rhys, Rich Farmbrough, Rifleman 82, Rintintin, Rjwilmsi, Rmky87, RobHutten, Rror, Ruud Koot, Rvencio, S177, Schutz, Scilit, Scottzed, Seglea, Senator Palpatine, SexyGod, Shawnc, Shortliffe, Shubinator, Shyamal, Sjoerd de Vries, Smjc, Smoe, Spin2cool, Steinsky, Stewartadcock, Stinkbeard, Subhashis.behera, Supten, Surajbodi, Susurrus, Tapir Terrific, Tarcieri, Tdhoufek, Tellyaddict, Template namespace initialisation script, Terrace4, TestPilot, The New Mikemoral, Thenothing, Thermochap, Thomaswgc, Thorwald, Thumperward, Tim@, TimVickers, Tincup, Tmccrae, Tombadog, Tompw, Tpvipin, Tupeliano, Turnstep, Vanka5, VashiDonsk, Vasundhar, Vawter, Vegasprof, Venus Victorious, Veterinarian, Vietbio, Vina, Viriditas, W09110900, Walshga, Wavelength, Wieghardt, Wik, Wiki1forall, WilliamBonfieldFRS, Willking1979, Winhide, Wmahan, Woohookitty, Ymichel, Yoni-vL, Youssefsan, Zashaw, ZayZayEM, Zhuozhuo, Zoicon5, Zorozorozoro123, Zzuuzz, आशीष भटनागर, 499 anonymous edits

Image Sources, Licenses and Contributors

File:ADN animation.gif *Source*: http://en.wikipedia.org/w/index.php?title=File:ADN_animation.gif *License*: Public Domain *Contributors*: Aushulz, Bawolff, Brian0918, Kersti Nebelsiek, Magadan, Mattes, Origamiemensch, Stevenfruitsmaak, 3 anonymous edits

Image:DNA chemical structure.svg *Source*: http://en.wikipedia.org/w/index.php?title=File:DNA_chemical_structure.svg *License*: unknown *Contributors*: Madprime, Wickey, 1 anonymous edits

File:DNA orbit animated static thumb.png *Source*: http://en.wikipedia.org/w/index.php?title=File:DNA_orbit_animated_static_thumb.png *License*: GNU Free Documentation License *Contributors*: 84user adapting file originally uploaded by Richard Wheeler (Zephyris) at en.wikipedia

Image:GC DNA base pair.svg *Source*: http://en.wikipedia.org/w/index.php?title=File:GC_DNA_base_pair.svg *License*: Public Domain *Contributors*: User:Isilanes

Image:AT DNA base pair.svg *Source*: http://en.wikipedia.org/w/index.php?title=File:AT_DNA_base_pair.svg *License*: Public Domain *Contributors*: User:Isilanes

Image:A-DNA, B-DNA and Z-DNA.png *Source*: http://en.wikipedia.org/w/index.php?title=File:A-DNA,_B-DNA_and_Z-DNA.png *License*: GNU Free Documentation License *Contributors*: Original uploader was Richard Wheeler (Zephyris) at en.wikipedia

Image:Parallel telomere quadruple.png *Source*: http://en.wikipedia.org/w/index.php?title=File:Parallel_telomere_quadruple.png *License*: unknown *Contributors*: User:Splette

Image:Branch-dna.png *Source*: http://en.wikipedia.org/w/index.php?title=File:Branch-dna.png *License*: unknown *Contributors*: Peter K.

Image:Multi-branch-dna.png *Source*: http://en.wikipedia.org/w/index.php?title=File:Multi-branch-dna.png *License*: unknown *Contributors*: User:Peter K.

Image:Cytosine chemical structure.png *Source*: http://en.wikipedia.org/w/index.php?title=File:Cytosine_chemical_structure.png *License*: GNU Free Documentation License *Contributors*: BorisTM, Bryan Derksen, Cacycle, Edgar181, Engineer gena

Image:5-methylcytosine.png *Source*: http://en.wikipedia.org/w/index.php?title=File:5-methylcytosine.png *License*: unknown *Contributors*: EDUCA33E, Luigi Chiesa, Mysid

Image:Thymine chemical structure.png *Source*: http://en.wikipedia.org/w/index.php?title=File:Thymine_chemical_structure.png *License*: GNU Free Documentation License *Contributors*: Arrowsmaster, BorisTM, Bryan Derksen, Cacycle, Edgar181, Leyo

Image:Benzopyrene DNA adduct 1JDG.png *Source*: http://en.wikipedia.org/w/index.php?title=File:Benzopyrene_DNA_adduct_1JDG.png *License*: GNU Free Documentation License *Contributors*: Benjah-bmm27, Bstlee, 1 anonymous edits

Image:T7 RNA polymerase at work.png *Source*: http://en.wikipedia.org/w/index.php?title=File:T7_RNA_polymerase_at_work.png *License*: unknown *Contributors*: User:Splette

Image:DNA replication.svg *Source*: http://en.wikipedia.org/w/index.php?title=File:DNA_replication.svg *License*: Public Domain *Contributors*: user:LadyofHats

Image:Nucleosome 2.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:Nucleosome_2.jpg *License*: Public Domain *Contributors*: Original uploader was TimVickers at en.wikipedia

Image:Nucleosome (opposites attracts).JPG *Source*: http://en.wikipedia.org/w/index.php?title=File:Nucleosome_(opposites_attracts).JPG *License*: Public Domain *Contributors*: Illustration by David S. Goodsell of The Scripps Research Institute (see this site)

Image:Lambda repressor 1LMB.png *Source*: http://en.wikipedia.org/w/index.php?title=File:Lambda_repressor_1LMB.png *License*: GNU Free Documentation License *Contributors*: Original uploader was Zephyris at en.wikipedia

Image:EcoRV 1RVA.png *Source*: http://en.wikipedia.org/w/index.php?title=File:EcoRV_1RVA.png *License*: GNU Free Documentation License *Contributors*: Original uploader was Zephyris at en.wikipedia

Image:Holliday Junction cropped.png *Source*: http://en.wikipedia.org/w/index.php?title=File:Holliday_Junction_cropped.png *License*: GNU Free Documentation License *Contributors*: Original uploader was TimVickers at en.wikipedia

Image:Holliday junction coloured.png *Source*: http://en.wikipedia.org/w/index.php?title=File:Holliday_junction_coloured.png *License*: GNU Free Documentation License *Contributors*: Original uploader was Zephyris at en.wikipedia

Image:Chromosomal Recombination.svg *Source*: http://en.wikipedia.org/w/index.php?title=File:Chromosomal_Recombination.svg *License*: Creative Commons Attribution 2.5 *Contributors*: User:Gringer

Image:DNA nanostructures.png *Source*: http://en.wikipedia.org/w/index.php?title=File:DNA_nanostructures.png *License*: unknown *Contributors*: (Images were kindly provided by Thomas H. LaBean and Hao Yan.)

Image:JamesDWatson.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:JamesDWatson.jpg *License*: Public Domain *Contributors*: Edward, RP88, Shizhao, Stellatomailing, Vonvon, 3 anonymous edits

Image:Francis Crick.png *Source*: http://en.wikipedia.org/w/index.php?title=File:Francis_Crick.png *License*: unknown *Contributors*: Photo: Marc Lieberman

Image:FrancisHarryComptonCrick.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:FrancisHarryComptonCrick.jpg *License*: unknown *Contributors*: Bunzil, Million Moments, PDH

Image:Rosalind Franklin.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:Rosalind_Franklin.jpg *License*: Public Domain *Contributors*: unknown

Image:Raymond Gosling.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:Raymond_Gosling.jpg *License*: GNU Free Documentation License *Contributors*: User:Davidruben

Image:maurice_wilkins.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:Maurice_wilkins.jpg *License*: unknown *Contributors*: Anetode, PDH, Rvilbig

Image:Erwin Chargaff.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:Erwin_Chargaff.jpg *License*: Public Domain *Contributors*: unknown

Image:Rosalindfranklinsjokecard.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:Rosalindfranklinsjokecard.jpg *License*: unknown *Contributors*: Bci2, Martyman, Nitramrekcap, Rjm at sleepers, 2 anonymous edits

Image:Sarfus.DNABiochip.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:Sarfus.DNABiochip.jpg *License*: unknown *Contributors*: Nanolane

Image:Spinning DNA.gif *Source*: http://en.wikipedia.org/w/index.php?title=File:Spinning_DNA.gif *License*: Public Domain *Contributors*: USDA

File:Methanol.pdb.png *Source*: http://en.wikipedia.org/w/index.php?title=File:Methanol.pdb.png *License*: Creative Commons Attribution-Sharealike 2.5 *Contributors*: ALoopingIcon, Benjah-bmm27

File:DNA-fragment-3D-vdW.png *Source*: http://en.wikipedia.org/w/index.php?title=File:DNA-fragment-3D-vdW.png *License*: Public Domain *Contributors*: Benjah-bmm27

File:Simple harmonic oscillator.gif *Source*: http://en.wikipedia.org/w/index.php?title=File:Simple_harmonic_oscillator.gif *License*: Public Domain *Contributors*: User:Oleg Alexandrov

File:DNA chemical structure.svg *Source*: http://en.wikipedia.org/w/index.php?title=File:DNA_chemical_structure.svg *License*: unknown *Contributors*: Madprime, Wickey, 1 anonymous edits

File:Parallel telomere quadruple.png *Source*: http://en.wikipedia.org/w/index.php?title=File:Parallel_telomere_quadruple.png *License*: unknown *Contributors*: User:Splette

File:Four-way DNA junction.gif *Source*: http://en.wikipedia.org/w/index.php?title=File:Four-way_DNA_junction.gif *License*: Public Domain *Contributors*: Aushulz, Molatwork, Origamiemensch, TimVickers, 1 anonymous edits

File:DNA replication.svg *Source*: http://en.wikipedia.org/w/index.php?title=File:DNA_replication.svg *License*: Public Domain *Contributors*: user:LadyofHats

File:ABDNAxrgpj.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:ABDNAxrgpj.jpg *License*: GNU Free Documentation License *Contributors*: I.C. Baianu et al.

File:Plos VHL.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:Plos_VHL.jpg *License*: Creative Commons Attribution 2.5 *Contributors*: Akinom, Anniolek, Filip em, Thommiddleton

File:3D model hydrogen bonds in water.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:3D_model_hydrogen_bonds_in_water.jpg *License*: GNU Free Documentation License *Contributors*: User:snek01

Image:Methanol.pdb.png *Source*: http://en.wikipedia.org/w/index.php?title=File:Methanol.pdb.png *License*: Creative Commons Attribution-Sharealike 2.5 *Contributors*: ALoopingIcon, Benjah-bmm27

File:Bragg diffraction.png *Source*: http://en.wikipedia.org/w/index.php?title=File:Bragg_diffraction.png *License*: GNU General Public License *Contributors*: user:hadmack

File:DNA in water.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:DNA_in_water.jpg *License*: unknown *Contributors*: User:Bbkkk

File:X ray diffraction.png *Source*: http://en.wikipedia.org/w/index.php?title=File:X_ray_diffraction.png *License*: unknown *Contributors*: Thomas Splettstoesser

File:X Ray Diffractometer.JPG *Source*: http://en.wikipedia.org/w/index.php?title=File:X_Ray_Diffractometer.JPG *License*: GNU Free Documentation License *Contributors*: Ff02::3, Pieter Kuiper

File:SLAC detector edit1.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:SLAC_detector_edit1.jpg *License*: unknown *Contributors*: User:Mfield, User:Starwiz

File:ISIS exptal hall.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:ISIS_exptal_hall.jpg *License*: unknown *Contributors*: wurzeller

File:Dna-SNP.svg *Source*: http://en.wikipedia.org/w/index.php?title=File:Dna-SNP.svg *License*: unknown *Contributors*: User:Gringer

File:DNA Under electron microscope Image 3576B-PH.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:DNA_Under_electron_microscope_Image_3576B-PH.jpg *License*: unknown *Contributors*: Original uploader was SeanMack at en.wikipedia

File:DNA Model Crick-Watson.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:DNA_Model_Crick-Watson.jpg *License*: Public Domain *Contributors*: User:Alkivar

File:DNA labels.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:DNA_labels.jpg *License*: GNU Free Documentation License *Contributors*: User:Raul654

File:AT DNA base pair pt.svg *Source*: http://en.wikipedia.org/w/index.php?title=File:AT_DNA_base_pair_pt.svg *License*: Public Domain *Contributors*: User:Lijealso

File:A-B-Z-DNA Side View.png *Source*: http://en.wikipedia.org/w/index.php?title=File:A-B-Z-DNA_Side_View.png *License*: Public Domain *Contributors*: Original uploader was Thorwald at en.wikipedia

File:Museo Príncipe Felipe. ADN.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:Museo_Príncipe_Felipe._ADN.jpg *License*: Creative Commons Attribution-Sharealike 2.0 *Contributors*: Fernando

File:AGCT DNA mini.png *Source*: http://en.wikipedia.org/w/index.php?title=File:AGCT_DNA_mini.png *License*: unknown *Contributors*: Iquo

File:BU Bio5.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:BU_Bio5.jpg *License*: Creative Commons Attribution-Sharealike 2.0 *Contributors*: Original uploader was Elapied at fr.wikipedia

File:Circular DNA Supercoiling.png *Source*: http://en.wikipedia.org/w/index.php?title=File:Circular_DNA_Supercoiling.png *License*: GNU Free Documentation License *Contributors*: Richard Wheeler (Zephyris)

File:Rosalindfranklinsjokecard.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:Rosalindfranklinsjokecard.jpg *License*: unknown *Contributors*: Bci2, Martyman, Nitramrekcap, Rjm at sleepers, 2 anonymous edits

File:Genomics GTL Pictorial Program.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:Genomics_GTL_Pictorial_Program.jpg *License*: Public Domain *Contributors*: Mdd

File:RNA pol.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:RNA_pol.jpg *License*: Public Domain *Contributors*: InfoCan

File:Primase 3B39.png *Source*: http://en.wikipedia.org/w/index.php?title=File:Primase_3B39.png *License*: Public Domain *Contributors*: own work

File:DNA Repair.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:DNA_Repair.jpg *License*: Public Domain *Contributors*: Courtesy of Tom Ellenberger, Washington University School of Medicine in St. Louis.

File:MGMT+DNA 1T38.png *Source*: http://en.wikipedia.org/w/index.php?title=File:MGMT+DNA_1T38.png *License*: Public Domain *Contributors*: own work

File:DNA damaged by carcinogenic 2-aminofluorene AF .jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:DNA_damaged_by_carcinogenic_2-aminofluorene_AF_.jpg *License*: Public Domain *Contributors*: Brian E. Hingerty, Oak Ridge National Laboratory Suse Broyde, New York University Dinshaw J. Patel, Memorial Sloan Kettering Cancer Center

File:A-DNA orbit animated small.gif *Source*: http://en.wikipedia.org/w/index.php?title=File:A-DNA_orbit_animated_small.gif *License*: GNU Free Documentation License *Contributors*: User:Bstlee, User:Zephyris

File:Plasmid emNL.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:Plasmid_emNL.jpg *License*: GNU Free Documentation License *Contributors*: Denniss, Glenn, Rasbak

File:Chromatin chromosom.png *Source*: http://en.wikipedia.org/w/index.php?title=File:Chromatin_chromosom.png *License*: Public Domain *Contributors*: User:Magnus Manske

File:Chromosome.svg *Source*: http://en.wikipedia.org/w/index.php?title=File:Chromosome.svg *License*: unknown *Contributors*: User:Dietzel65, User:Magnus Manske, User:Tryphon

File:Chr2 orang human.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:Chr2_orang_human.jpg *License*: Creative Commons Attribution-Sharealike 2.5 *Contributors*: Verena Schubel, Stefan Müller, Department Biologie der Ludwig-Maximilians-Universität München.

File:3D-SIM-3 Prophase 3 color.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:3D-SIM-3_Prophase_3_color.jpg *License*: Creative Commons Attribution-Sharealike 3.0 *Contributors*: Lothar Schermelleh

File:Chromosome2 merge.png *Source*: http://en.wikipedia.org/w/index.php?title=File:Chromosome2_merge.png *License*: Public Domain *Contributors*: Original uploader was Evercat at en.wikipedia

File:Transkription Translation 01.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:Transkription_Translation_01.jpg *License*: Public Domain *Contributors*: User:Kuebi

File:RibosomaleTranskriptionsEinheit.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:RibosomaleTranskriptionsEinheit.jpg *License*: GNU Free Documentation License *Contributors*: User:Merops

File:Chromosome Conformation Capture Technology.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:Chromosome_Conformation_Capture_Technology.jpg *License*: Public Domain *Contributors*: User:Kangyun1985

File:Mitochondrial DNA and diseases.png *Source*: http://en.wikipedia.org/w/index.php?title=File:Mitochondrial_DNA_and_diseases.png *License*: unknown *Contributors*: User:XXXL1986

File:PCR.svg *Source*: http://en.wikipedia.org/w/index.php?title=File:PCR.svg *License*: unknown *Contributors*: User:Madprime

File:Pcr gel.png *Source*: http://en.wikipedia.org/w/index.php?title=File:Pcr_gel.png *License*: GNU Free Documentation License *Contributors*: Habj, Ies, PatríciaR, Retama, Saperaud

File:DNA nanostructures.png *Source*: http://en.wikipedia.org/w/index.php?title=File:DNA_nanostructures.png *License*: unknown *Contributors*: (Images were kindly provided by Thomas H. LaBean and Hao Yan.)

File:SFP discovery principle.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:SFP_discovery_principle.jpg *License*: unknown *Contributors*: User:Agbiotec

File:Cdnaarray.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:Cdnaarray.jpg *License*: unknown *Contributors*: Mangapoco

File:Expression of Human Wild-Type and P239S Mutant Palladin.png *Source*: http://en.wikipedia.org/w/index.php?title=File:Expression_of_Human_Wild-Type_and_P239S_Mutant_Palladin.png *License*: unknown *Contributors*: see above

File:Random genetic drift chart.png *Source*: http://en.wikipedia.org/w/index.php?title=File:Random_genetic_drift_chart.png *License*: unknown *Contributors*: User:Professor marginalia

File:Co-dominance Rhododendron.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:Co-dominance_Rhododendron.jpg *License*: Creative Commons Attribution 2.0 *Contributors*: Ayacop, Cillas, FlickrLickr, FlickreviewR, Horcha, Kanonkas, Kevmin, MPF, Para

File:DNA_nanostructures.png *Source*: http://en.wikipedia.org/w/index.php?title=File:DNA_nanostructures.png *License*: unknown *Contributors*: (Images were kindly provided by Thomas H. LaBean and Hao Yan.)

File:Holliday junction coloured.png *Source*: http://en.wikipedia.org/w/index.php?title=File:Holliday_junction_coloured.png *License*: GNU Free Documentation License *Contributors*: Original uploader was Zephyris at en.wikipedia

File:Holliday Junction cropped.png *Source*: http://en.wikipedia.org/w/index.php?title=File:Holliday_Junction_cropped.png *License*: GNU Free Documentation License *Contributors*: Original uploader was TimVickers at en.wikipedia

File:Atomic force microscope by Zureks.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:Atomic_force_microscope_by_Zureks.jpg *License*: unknown *Contributors*: User:Zureks

File:Atomic force microscope block diagram.png *Source*: http://en.wikipedia.org/w/index.php?title=File:Atomic_force_microscope_block_diagram.png *License*: Public Domain *Contributors*: Original uploader was Askewmind at en.wikipedia

File:AFM view of sodium chloride.gif *Source*: http://en.wikipedia.org/w/index.php?title=File:AFM_view_of_sodium_chloride.gif *License*: Public Domain *Contributors*: Courtesy of prof. Ernst Meyer, university of Basel

File:Single-Molecule-Under-Water-AFM-Tapping-Mode.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:Single-Molecule-Under-Water-AFM-Tapping-Mode.jpg *License*: unknown *Contributors*: User:Yurko

File:AFMimageRoughGlass20x20.png *Source*: http://en.wikipedia.org/w/index.php?title=File:AFMimageRoughGlass20x20.png *License*: Public Domain *Contributors*: Chych

File:Maldi informatics figure 6.JPG *Source*: http://en.wikipedia.org/w/index.php?title=File:Maldi_informatics_figure_6.JPG *License*: Public Domain *Contributors*: Rbeavis

File:Stokes shift.png *Source*: http://en.wikipedia.org/w/index.php?title=File:Stokes_shift.png *License*: unknown *Contributors*: User:Mykhal

File:CARS Scheme.svg *Source*: http://en.wikipedia.org/w/index.php?title=File:CARS_Scheme.svg *License*: unknown *Contributors*: Onno Gabriel

File:HyperspectralCube.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:HyperspectralCube.jpg *License*: Public Domain *Contributors*: Dr. Nicholas M. Short, Sr.

File:MultispectralComparedToHyperspectral.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:MultispectralComparedToHyperspectral.jpg *License*: Public Domain *Contributors*: Dr. Nicholas M. Short, Sr.

File:Confocalprinciple.svg *Source*: http://en.wikipedia.org/w/index.php?title=File:Confocalprinciple.svg *License*: GNU Free Documentation License *Contributors*: Danh

File:3D-SIM-1 NPC Confocal vs 3D-SIM detail.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:3D-SIM-1_NPC_Confocal_vs_3D-SIM_detail.jpg *License*: Creative Commons Attribution-Sharealike 3.0 *Contributors*: Changes in layout by the uploader. Only the creator of the original (Lothar Schermelleh) should be credited.

File:Tirfm.svg *Source*: http://en.wikipedia.org/w/index.php?title=File:Tirfm.svg *License*: Public Domain *Contributors*: Dawid Kulik

File:Inverted microscope.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:Inverted_microscope.jpg *License*: unknown *Contributors*: Nuno Nogueira (Nmnogueira) Original uploader was Nmnogueira at en.wikipedia

File:Fluorescence microscop.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:Fluorescence_microscop.jpg *License*: unknown *Contributors*: Masur

File:Microscope And Digital Camera.JPG *Source*: http://en.wikipedia.org/w/index.php?title=File:Microscope_And_Digital_Camera.JPG *License*: GNU Free Documentation License *Contributors*: User:Zephyris

File:FluorescenceFilters 2008-09-28.svg *Source*: http://en.wikipedia.org/w/index.php?title=File:FluorescenceFilters_2008-09-28.svg *License*: unknown *Contributors*: User:Mastermolch

File:FluorescentCells.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:FluorescentCells.jpg *License*: Public Domain *Contributors*: DO11.10, Emijrp, NEON ja, Origamiemensch, Splette, Tolanor, 5 anonymous edits

File:Yeast membrane proteins.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:Yeast_membrane_proteins.jpg *License*: unknown *Contributors*: User:Masur

File:S cerevisiae septins.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:S_cerevisiae_septins.jpg *License*: Public Domain *Contributors*: Spitfire ch, Philippsen Lab, Biozentrum Basel

File:Dividing Cell Fluorescence.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:Dividing_Cell_Fluorescence.jpg *License*: unknown *Contributors*: Will-moore-dundee

File:HeLa Hoechst 33258.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:HeLa_Hoechst_33258.jpg *License*: Public Domain *Contributors*: TenOfAllTrades

File:FISH 13 21.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:FISH_13_21.jpg *License*: Public Domain *Contributors*: Gregor1976

File:300px-Anaphase-fluorescent.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:300px-Anaphase-fluorescent.jpg *License*: GNU Free Documentation License *Contributors*: New York State Department of Health

File:Bloodcell sun flares pathology.jpeg *Source*: http://en.wikipedia.org/w/index.php?title=File:Bloodcell_sun_flares_pathology.jpeg *License*: Public Domain *Contributors*: Birindand, Karelj, NEON ja, 1 anonymous edits

File:Carboxysome 3 images.png *Source*: http://en.wikipedia.org/w/index.php?title=File:Carboxysome_3_images.png *License*: Creative Commons Attribution 3.0 *Contributors*: Prof. Todd O. Yeates, UCLA Dept. of Chem. and Biochem.

Image:protein pattern analyzer.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:Protein_pattern_analyzer.jpg *License*: Public Domain *Contributors*: Athaenara, Emesee, 1 anonymous edits

Image:A thaliana metabolic network.png *Source*: http://en.wikipedia.org/w/index.php?title=File:A_thaliana_metabolic_network.png *License*: GNU Free Documentation License *Contributors*: Original uploader was TimVickers at en.wikipedia

Image:Metabolic Network Model for Escherichia coli.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:Metabolic_Network_Model_for_Escherichia_coli.jpg *License*: Public Domain *Contributors*: Mdd

Image:Metabolism 790px partly labeled.png *Source*: http://en.wikipedia.org/w/index.php?title=File:Metabolism_790px_partly_labeled.png *License*: GNU Free Documentation License *Contributors*: See below

Image:Gtk-dialog-info.svg *Source*: http://en.wikipedia.org/w/index.php?title=File:Gtk-dialog-info.svg *License*: GNU Lesser General Public License *Contributors*: David Vignoni

Image:Metabolism 790px.png *Source*: http://en.wikipedia.org/w/index.php?title=File:Metabolism_790px.png *License*: GNU Free Documentation License *Contributors*: See below

Image:Cell cycle bifurcation diagram.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:Cell_cycle_bifurcation_diagram.jpg *License*: unknown *Contributors*: User:Squidonius

Image:Genomics GTL Pictorial Program.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:Genomics_GTL_Pictorial_Program.jpg *License*: Public Domain *Contributors*: Mdd

Image:Signal transduction v1.png *Source*: http://en.wikipedia.org/w/index.php?title=File:Signal_transduction_v1.png *License*: GNU Free Documentation License *Contributors*: Original uploader was Roadnottaken at en.wikipedia

Image:Insulincrystals.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:Insulincrystals.jpg *License*: Public Domain *Contributors*: Chrumps, Jurema Oliveira, Photohound

Image:16thCenturyBrewer.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:16thCenturyBrewer.jpg *License*: Public Domain *Contributors*: User SilkTork on en.wikipedia

Image:99341.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:99341.jpg *License*: Public Domain *Contributors*: Scott Bauer

Image:Microarray2.gif *Source*: http://en.wikipedia.org/w/index.php?title=File:Microarray2.gif *License*: Public Domain *Contributors*: Original uploader was Paphrag at en.wikipedia

Image:InsulinHexamer.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:InsulinHexamer.jpg *License*: Creative Commons Attribution 2.5 *Contributors*: Original uploader was Takometer at en.wikipedia

Image:Gel electrophoresis 2.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:Gel_electrophoresis_2.jpg *License*: Creative Commons Attribution-Sharealike 2.0 *Contributors*: User:Mnolf

Image:E coli at 10000x, original.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:E_coli_at_10000x,_original.jpg *License*: Public Domain *Contributors*: Photo by Eric Erbe, digital colorization by Christopher Pooley, both of USDA, ARS, EMU.

Image:Gene therapy.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:Gene_therapy.jpg *License*: unknown *Contributors*: Ies, Krimpet, Llull, TimVickers, Ysangkok

Image:Dna-split.png *Source*: http://en.wikipedia.org/w/index.php?title=File:Dna-split.png *License*: Public Domain *Contributors*: Broadbeer, Conscious, Dietzel65, LadyofHats, Madprime, Magnus Manske, Niki K, Pixeltoo, Roybb95, Samulili, Teetaweepo, 1 anonymous edits

Image:Genome viewer screenshot small.png *Source*: http://en.wikipedia.org/w/index.php?title=File:Genome_viewer_screenshot_small.png *License*: Public Domain *Contributors*:

License

Version 1.2, November 2002

0.PREAMBLE

The purpose of this License is to make a manual, textbook, or other functional and useful document "free" in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondarily, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.
This License is a kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.
We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

1.APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The "Document", below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as "you". You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.
A "Modified Version" of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.
A "Secondary Section" is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.
The "Invariant Sections" are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.
The "Cover Texts" are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.
A "Transparent" copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not "Transparent" is called "Opaque".
Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.
The "Title Page" means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, "Title Page" means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.
A section "Entitled XYZ" means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as "Acknowledgements", "Dedications", "Endorsements", or "History".) To "Preserve the Title" of such a section when you modify the Document means that it remains a section "Entitled XYZ" according to this definition.
The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

2.VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.
You may also lend copies, under the same conditions stated above, and you may publicly display copies.

3.COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.
If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.
If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.
It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

4.MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

1. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.
2. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.
3. State on the Title page the name of the publisher of the Modified Version, as the publisher.
4. Preserve all the copyright notices of the Document.
5. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.

6. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.
7. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.
8. Include an unaltered copy of this License.
9. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.
10. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.
11. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.
12. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.
13. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version.
14. Do not retitle any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section.
15. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties--for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

5.COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled "History" in the various original documents, forming one section Entitled "History"; likewise combine any sections Entitled "Acknowledgements", and any sections Entitled "Dedications". You must delete all sections Entitled "Endorsements."

6.COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

7.AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an "aggregate" if the copyright resulting from the compilation is not used to limit the legal rights of the compilation's users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document's Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

8.TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled "Acknowledgements", "Dedications", or "History", the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

9.TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided for under this License. Any other attempt to copy, modify, sublicense or distribute the Document is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

10.FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See http://www.gnu.org/copyleft/.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation.

How to use this License for your documents

To use this License in a document you have written, include a copy of the License in the document and put the following copyright and license notices just after the title page:

> Copyright (c) YEAR YOUR NAME.
>
> Permission is granted to copy, distribute and/or modify this document
>
> under the terms of the GNU Free Documentation License, Version 1.2
>
> or any later version published by the Free Software Foundation;
>
> with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts.
>
> A copy of the license is included in the section entitled "GNU
>
> Free Documentation License".

If you have Invariant Sections, Front-Cover Texts and Back-Cover Texts, replace the "with...Texts." line with this:

> with the Invariant Sections being LIST THEIR TITLES, with the
>
> Front-Cover Texts being LIST, and with the Back-Cover Texts being LIST.

If you have Invariant Sections without Cover Texts, or some other combination of the three, merge those two alternatives to suit the situation.

If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free software.